

METABOLIC ENGINEERING OF ISOFLAVONOID BIOSYNTHESIS IN TOBACCO AND WHITE CLOVER

A thesis submitted in partial fulfilment
of the requirements for the Degree of
Doctor of Philosophy in Plant Biotechnology
in the University of Canterbury

Benjamin Konrad Franzmayr

University of Canterbury

2011

Dedication

This thesis is dedicated to my family. Demelza, Katja, Linnéa, Orion and Natalya – I love you very much and thank you for the sacrifices you have made, knowingly or not, to allow this research and thesis to be completed.

Abstract

Isoflavonoids are a class of plant secondary metabolites which have multiple biological roles in plants as pest feeding deterrents, phytoalexins and signals to rhizobial microbes. Some isoflavonoids, or their breakdown products, are estrogenic when ingested by animals, and pastures with high levels of the isoflavonoid formononetin can cause sterility in ewes. White clover has low levels of isoflavonoids and is susceptible to pests like the clover root weevil. The overall aim of this project was to test whether isoflavonoids could be manipulated in white clover through metabolic engineering.

The genes of the key isoflavonoid biosynthesis enzymes have been cloned from a range of legumes and three major genes, chalcone reductase (CHR), isoflavone synthase (IFS) and isoflavonoid O-methyltransferase (IOMT), were cloned from white clover in this study. The white clover IFS2_12 gene was expressed in transgenic tobacco. Genistein, an isoflavonoid that is not naturally present in tobacco, was detected in the IFS-expressing tobacco, thus confirming the functionality of the IFS2_12 gene. Tobacco plants were transformed with ANT1, a transcription factor that induces the production of anthocyanins that share precursors with the isoflavonoid biosynthesis pathway. When IFS was expressed in red tobacco leaves, where anthocyanin biosynthesis was occurring, the levels of genistein were greater than in anthocyanin-free green leaves.

White clover was transformed to overexpress the cloned IFS2_12 gene and some transformants had greater levels of IFS gene expression, up to 12.9 times the average wild type level. However, these transformants did not produce formononetin levels greater than the wild-type. A gene fusion of alfalfa chalcone isomerase (CHI), which produces the precursors naringenin and liquiritigenin, and soybean IFS, which converts the precursors to genistein and daidzein, respectively, was received from the Noble Foundation. Transgenic white clover plants expressing IFS/CHI were produced using a novel method that also regenerated wild-type clones of the transgenic plants. When compared with their wild-type clones, two IFS/CHI transformants produced higher levels of formononetin, thus supporting the suggestion that isoflavonoid levels can be increased in white clover through overexpression of isoflavonoid biosynthesis genes.

Acknowledgements

I sincerely thank my supervisors for all of their help and guidance along the way, in particular Paula Jameson for your keen thesis-checking skills, optimism and your warm and understanding nature and Susanne Rasmussen for conceiving of this PhD project, guidance on the research and for teaching me your biochemistry knowledge.

Thank you, Chris Jones, for supporting and encouraging me throughout my PhD. Thanks Andrew Griffiths for your good advice and selfless support despite losing me as your technician during this project.

I have been fortunate to have received help whenever I asked for it from my many wonderful colleagues at AgResearch Grasslands. Thank you to Kerry Hancock for help with techniques and consumables, Margaret Greig for making media and looking after plants, Dorothy Maher for help and materials for white clover transformations, Nick Ellison for lab techniques, Vern Collette for help with Vector NTI, John Ford and Tony Dunn for plant material, Chris Mercer for advice on plant propagation, Geoff Lane and Brian Tapper for discussions on plant biochemistry, Bill Rumball for discussions on red clover isoflavonoids, Jim Crush for sharing his CRW trial results, Kim Richardson for help with transformation vectors and techniques, Nick Roberts for discussions on model plants and research directions, Ruth Vlaming for creating the pRS vectors, Karl Fraser and Dan Hughes for running LCMS analyses and teaching me about ion traps, Zulfi Jahufer and Chikako van Koten for guidance on statistical analysis and Jean-Hugues (Jimmy) Hatier, Nick Albert, Andrew Griffiths and Tina Voisey for providing valuable feedback on thesis drafts. I thank the following people for their general help and support: Alicia Scott, Anne Allan, Craig Anderson, Dale Rosvall, Linda Johnson, Rainer Hofmann, Rachael Sheridan, Tash Forester and Tore Rayner.

For the people who have helped me and who I have inadvertently left out: Thank you!

Thank you to my parents in law, Brenda and Ernest, for your constant support of our family, it has helped a lot in these challenging years. Thanks to my Mama, Cisca, for taking care of me when I was staying in Christchurch. Thanks to my Papa, Clemens, for your support and advice. Thank you to my whanau for your encouragement and help.

This research was supported by an Enterprise Scholarship administered by the Tertiary Education Commission and the University of Canterbury. The scholarship was paid by the Foundation for Research, Science and Technology and by the Pastoral Genomics Consortium (Fonterra, Beef + Lamb New Zealand, AgResearch, DairyNZ and DEEResearch) which was set up to use molecular biology for the genetic improvement of forage plants. AgResearch has further supported me financially and supplied consumables and access to scientific equipment, expertise and other resources.

I acknowledge Li Tian and Richard Dixon of the Samuel Roberts Noble Foundation (Ardmore, OK, USA) for the use of the IFS/CHI fusion gene.

Table of Contents

Abstract.....	i
Acknowledgements	ii
Table of Contents.....	iv
List of Figures.....	x
List of Tables	xiv
Abbreviations	xv
1 Introduction	1
1.1 Background.....	1
1.2 White clover and its importance as a pasture plant.	2
1.3 Isoflavonoids.....	4
1.4 Plant protection and signalling	6
1.4.1 Isoflavonoids as plant protectors.....	6
1.4.2 Isoflavonoids as plant signals.....	9
1.4.3 Effects on animal health	9
1.5 Isoflavonoid biosynthesis	10
1.5.1 Chalcone reductase and chalcone isomerase.	12
1.5.2 Isoflavone synthase.....	15
1.5.3 Isoflavone O-methyltransferase (IOMT)	18

1.6	Metabolic engineering of isoflavonoids.....	19
1.7	Project Hypotheses, Aims and Objectives	23
2	Isolation and analysis of isoflavonoid biosynthesis genes from white clover.	25
2.1	Introduction.....	25
2.2	Materials and Methods	26
2.2.1	Locating genes in public literature, databases and the Pastoral Genomics white clover database	26
2.2.2	Contig assembly of Gene Thresher clones.....	26
2.2.3	Structural analysis of genes	26
2.2.4	Amplifying and cloning gene sequences from white clover DNA.	26
2.2.5	Standard PCR protocol	29
2.2.6	Triple Master PCR.....	29
2.2.7	Pfx 50 PCR	29
2.2.8	Agarose gel electrophoresis.....	30
2.2.9	Extracting DNA from agarose gels	30
2.2.10	Cloning of genes	30
2.2.11	Gateway BP and LR clonase reactions	31
2.2.12	Bacterial transformation	31
2.2.13	Testing plasmid for insert	32
2.2.14	Plasmid DNA extraction.....	32

2.2.15	Sequencing reaction	32
2.2.16	Plant materials	33
2.3	Results	34
2.3.1	Publicly available CHR, IFS and IOMT sequences	34
2.3.2	CHR, IFS and IOMT from the Pastoral Genomics white clover database and contig assembly of Gene Thresher clones	35
2.3.3	Structural analysis of genes	37
2.3.1	Amplifying and cloning of CHR, IFS and IOMT sequences from white clover DNA.	39
2.3.2	Analysis of cloned genes.	44
2.3.3	Chalcone reductase	44
2.3.4	Isoflavone synthase.....	47
2.3.5	Isoflavone O-methyltransferase	50
2.3.6	Cloning of ANT1 gene	53
2.3.7	Isoflavone synthase/ chalcone isomerase.....	53
2.4	Discussion	53
2.4.1	CHR.....	53
2.4.2	IFS	57
2.4.3	IOMT	58
2.4.4	ANT1	59

2.4.5	Summary	59
3	Expression of IFS in transgenic tobacco plants.....	61
3.1	Introduction.....	61
3.2	Materials and Methods	64
3.2.1	Plant materials.....	64
3.2.2	Transformation vectors.....	64
3.2.3	Transformation and culture of <i>Agrobacterium tumefaciens</i>	65
3.2.4	Stable transformation of <i>Nicotiana tabacum</i>	65
3.2.5	Plant DNA extraction	66
3.2.6	PCR confirmation of tobacco transformation.....	67
3.2.7	XNAP PCR	68
3.2.8	Transient expression of transgenes by leaf infiltration of <i>Nicotiana tabacum</i>	69
3.2.9	Transient expression experiments	70
3.2.10	Crossing of stable tobacco transformants.....	71
3.2.11	Testing of GFP and GUS expression	72
3.2.12	Isoflavonoid extraction	72
3.2.13	Quantification of isoflavonoids using high performance liquid chromatography (HPLC)	73
3.2.14	Liquid chromatography with mass spectrometric analysis (LCMS).....	74

3.3	Results	76
3.3.1	Stable tobacco transformants expressing isoflavonoid biosynthesis genes .	76
3.3.2	Stable tobacco transformants carrying ANT1	77
3.3.3	Transient expression in tobacco through leaf infiltration	81
3.3.4	Isoflavonoid content in infiltration experiments determined by LCMS	88
3.3.5	Isoflavonoid content in the progeny of stable tobacco transformant crosses.	91
3.4	Discussion	94
4	Transgenic white clover	100
4.1	Introduction	100
4.2	Materials and Methods	100
4.2.1	Plant materials	100
4.2.2	White clover transformation	102
4.2.3	Regeneration of transformed meristem through tissue culture	107
4.2.4	PCR confirmation of transformation	108
4.2.5	Extraction and quantification of RNA	102
4.2.6	Quantitative PCR of IFS cDNA	103
4.2.7	Determining natural isoflavonoid content and IFS gene expression in white and red clover	101
4.2.8	Determination of isoflavonoid levels in transgenic white clover	109
4.3	Results	109

4.3.1	Natural isoflavonoid levels and variability in white and red clover	109
4.3.2	Natural isoflavonoid content and IFS gene expression in white and red clover	115
4.3.3	Transformation and regeneration of white clover plants.....	116
4.3.4	Isoflavonoid content and IFS gene expression in Tr-IFS2_12 transgenic white clover	117
4.3.5	Isoflavonoid content of Tr-IFS/CHI transgenic white clover	120
4.4	Discussion	121
5	General Discussion	126
	References	132
	Appendix 1	147
	Appendix 2	150
	Appendix 3	153
	Appendix 4	154

List of Figures

Figure 1.1	Illustration of white clover.	3
Figure 1.2	Structures of the 13 known classes of isoflavonoids.	5
Figure 1.3	Outline of the isoflavonoid biosynthesis pathway in legumes.	11
Figure 1.4	Reaction catalysed by chalcone isomerase.	14
Figure 1.5	Production of genistein and daidzein by 2-hydroxy isoflavone synthase	16
Figure 2.1	Contigs of white clover genomic sequences from a Gene-Thresher database with high similarity to CHR, IFS and IOMT search queries.	36
Figure 2.2	Feature maps of CHR, IFS and IOMT contigs.	38
Figure 2.3	Products of PCR number 1 - CHR, IFS and IOMT from white clover	39
Figure 2.4	Triple Master PCR of genomic and cDNA of CHR, IFS and IOMT..	40
Figure 2.5	Standard PCR of genomic IFS and IOMT.	40
Figure 2.6	Standard PCR of IFS.	41
Figure 2.7	Triple Master PCR of IFS.	41
Figure 2.8	Standard PCR of CHR.	42
Figure 2.9	Pfx50 PCR of IOMT	42
Figure 2.10	Illustration of putative SNPs in CHRc clones from one unknown ‘Sustain’ genotype that allow the grouping of the sequences into five variants.	45

Figure 2.11	Phylogenetic tree of the amino acid sequence of representative sequences from nine CHR sequence variants from white clover and the functionally confirmed CHR P10 from soybean.	46
Figure 2.12	Alignment of amino acid sequence of <i>M. sativa</i> CHR12 and representative sequences from nine CHR sequence variants from white clover, showing amino acid residues of the ‘catalytic tetrad’ that are essential for CHR functioning.	47
Figure 2.13	Phylogenetic tree of the amino acid sequence of representative sequences from 13 putative IFS sequence variants from white clover and five functionally confirmed public IFS genes.	48
Figure 2.14	Phylogenetic tree of the amino acid sequence of representative sequences from 13 IFS sequence variants from white clover and 20 IFS genes from Jung et al. (2000).....	49
Figure 2.15	Alignment of <i>G. echinata</i> CYP93C2 and amino acid sequence of representative sequences from 13 putative IFS sequence variants from white clover showing amino acid residues that are essential for IFS functioning.	50
Figure 2.16	Phylogenetic tree of the amino acid sequence of representative sequences from eight IOMT sequence variants from white clover and nine functionally confirmed public IOMT genes.....	51
Figure 2.17	Alignment of MsI7OMT and IOMTg10, 12 and 22 showing amino acid residues that shape the enzymatic reaction pocket and are involved in SAM-binding and catalysis	52
Figure 2.18	PCR of ANT1	53
Figure 3.1	The structures of four isoflavone aglycones and their MS-MS fragment pathway.....	75
Figure 3.2	Test of Nt-IFS2_12 plants for presence of IFS2_12 transgene.	76

Figure 3.3	Test of Nt-IFS/CHI plants for presence of IFS/CHI transgene.	77
Figure 3.4	Test of Nt-ANT1 plants for presence of ANT1 transgene.	77
Figure 3.5	Nt-ANT1 regenerants showing the diversity of colour phenotypes.....	79
Figure 3.6	Nt-ANT1 regenerants showing phenotypes of both green and coloured areas in the leaves.	79
Figure 3.7	Phenotype of bronze tobacco Nt-ANT1 transformants.	80
Figure 3.8	Leaf infiltration in young tobacco leaves.....	81
Figure 3.9	GFP expression in leaves infiltrated with pRS13	82
Figure 3.10	GFP expression in Nt-ANT1 transgenic and wild type tobacco leaves infiltrated with pRS13.	83
Figure 3.11	Serial dilution of <i>A. tumefaciens</i> leaf infiltrates.....	85
Figure 3.12	Serial dilution of ANT1 <i>A. tumefaciens</i> leaf infiltrate in wild-type tobacco plants.	86
Figure 3.13	Nt-ANT1 leaves infiltrated without or with <i>A. tumefaciens</i> carrying IFS/CHI.	87
Figure 3.14	LCMS spectra of leaf tissue isoflavonoid extract from Nt-IFS/CHI transformant 5P02034 infiltrated with ANT1.	89
Figure 3.15	Isoflavonoid content of tobacco plants with combinations of stable and infiltrated IFS, ANT1 and negative control (pHZbar) transgenes.	90
Figure 3.16	Mass spectra of three peaks containing MS2 fragments of <i>m/z</i> 271.	91
Figure 3.17	Genistin content of seedlings from crossings of Nt-ANT1, Nt-IFS2_12 and Nt-IFS/CHI plants.	93

Figure 4.1	IFS qPCR standard curve and trendline.....	104
Figure 4.2	‘Virtual TLC’ visualisation of 50 HPLC chromatograms consisting of 10 replicates of five clover cultivars or genotypes.....	111
Figure 4.3	Average natural isoflavonoid content of red clover cultivars ‘Pawera’ and ‘G27’ and white clover cultivar ‘Sustain’ and clones of genotypes 21125 and 20161/21.....	113
Figure 4.4	Formononetin and biochanin A levels of four red clover cultivars in leaf, petiole and root tissues.	114
Figure 4.5	Isoflavonoid content and IFS gene expression in wild type red and white clover plants.....	115
Figure 4.6	PCR tests for the transgene in white clover IFS2_12 and IFS/CHI transformants.	117
Figure 4.7	Formononetin content of the leaves and roots, and expression of IFS in the leaves, of Tr-IFS2_12 regenerants.....	119
Figure 4.8	Formononetin glucoside malonate content of Tr-IFS/CHI transgenic plants and their wild-type clones.	120

List of Tables

Table 2.1	Summary of plant genotypes used.....	34
Table 2.2	A summary of CHR, IFS and IOMT genes cloned.....	43
Table 3.1	Primers used to test for transgenes in tobacco.....	68

Abbreviations

μl	microlitre
aa	amino acid
AMM	ammonium glufosinate (DL-phosphinothricin)
bp	basepair
CaMV 35S	cauliflower mosaic virus 35S promoter
cDNA	complementary DNA
CHI	chalcone isomerase
CHR	chalcone reductase
CRC	chimeric transcription factor containing maize C1 and R sequences
CRW	clover root weevil
d	day
DW	dry weight
DNA	deoxyribonucleic acid
EST	expressed sequence tag
FW	fresh weight
g	gram
<i>g</i>	gravity
gDNA	genomic DNA
GFP	green fluorescent protein
GUS	β-glucuronidase
h	hour
HPLC	high performance liquid chromatography
IFS	isoflavone synthase
IOMT	isoflavone O-methyltransferase
L	litre
LCMS	liquid chromatography mass spectrometry
min	minute
ml	millilitre
mRNA	messenger RNA
MS	mass spectrometry

Nt	<i>Nicotiana tabacum</i>
OCS	octopine synthase
ORF	open reading frame
qPCR	quantitative polymerase chain reaction.
RNA	ribonucleic acid
rpm	rotations per minute
s	second
t-DNA	transferred DNA
TIM	timentin (ticarcillin)
Tr	<i>Trifolium repens</i>
UV	ultraviolet

1 Introduction

1.1 Background

White clover is a valuable component of temperate pastures, providing nitrogen fixation to fertilise the soil and high quality nutrition to grazing livestock. However, white clover productivity is reduced by diseases and pests such as *Sitona lepidus*, the clover root weevil (CRW), which has recently been introduced to New Zealand. To maintain productivity, resistant or tolerant germplasm needs to be developed.

Feeding studies suggest that the isoflavonoids formononetin and biochanin A deter both CRW larvae which feed on the roots, and adults which feed on the leaves, of red clover (Gerard et al., 2005; Murray et al., 2007). White clover contains levels of formononetin and biochanin A that are too low to deter CRW feeding. A high-formononetin or biochanin A white clover would be expected to resist CRW damage but formononetin is a phytoestrogen and reduces fertility in sheep if they consume a large quantity of this isoflavonoid. Ewes grazing on some *Trifolium subterraneum* (subterranean clover) varieties in Western Australia had greatly reduced fertility and this was ascribed to the high formononetin content of the clover (Beck, 1964). Formononetin is demethylated and metabolised to equol by microorganisms in the animal digestive system and this estrogenic compound can build up to relatively high levels in the animal. Since biochanin A is only mildly estrogenic and is also an insect feeding deterrent, it may be useful to increase it in white clover to produce CRW resistance without causing potential problems with animal fertility.

The isoflavonoids genistein and daidzein, when exuded by the roots, are important signals to *Rhizobium sp.* bacteria for the establishment of root nodules where the bacteria reside and fix atmospheric nitrogen in a symbiotic relationship with the plant (Subramanian et al., 2006). This addition of nitrogen to a pasture is the most important role of white clover in a mixed sward and lifts the productivity of the pasture over that of unfertilised grass monocultures (Caradus et al. 1996).

The key isoflavonoid biosynthesis genes have been cloned in recent years and demonstrated to be functional when expressed as transgenes in other plants. This allows the

overexpression of the genes to increase isoflavonoid levels in legumes such as alfalfa (Deavours and Dixon, 2005), to increase phytoalexin levels, or in plants that do not produce any isoflavonoids, such as rice (Sreevidja et al., 2006), as a first step to engineering root nodulation and nitrogen fixation in this important food plant.

Given the diverse biological activity of isoflavonoids in plants and animals, in a white clover plant that was metabolically engineered to produce optimal isoflavonoid levels, the concentration, location and types of isoflavonoids produced would need to be carefully considered to balance the advantages of insect resistance and root nodulation with the potential disadvantages of estrogenicity in animals.

The main aim of this thesis was to explore whether it is possible to produce high levels of isoflavonoids in white clover using a metabolic engineering approach focused on the overexpression of an isoflavone synthase gene.

1.2 White clover and its importance as a pasture species

White clover (*Trifolium repens*) is a trifoliate perennial legume that grows in grasslands and multiplies both by seed and vegetatively through the growth of stolons along the ground (Figure 1.1). The main structural component of the plant is the stolon which consists of nodes and internodes. Each node produces one trifoliate leaf, a lateral bud and two nodal root buds. The root buds will grow if they come into contact with moist soil shortly after they emerge. Branching occurs from the lateral bud at the base of each leaf. At the growing tip of each stolon is the apical bud where leaves, lateral buds and flowers form. Branching is stimulated by light and the formation of nodal roots. The more branches that form, the more growing tips there are to produce further laterals, leaves and flowers. Over time, older parts of the stolon die off in the centre of the plant and the nodal root system then supports the new tips, resulting in fragmentation of the plant (Caradus et al., 1995).



Figure 1.1 Illustration of white clover. 1 trifoliate leaf, 2 unopened floret, 3 open floret, 4 fertilised floret, 5 flower head with all florets fertilised and dropped, 6 seed pod, 7 seed. Reproduced from Thomé (1885).

White clover forms a symbiotic relationship with *Rhizobia* bacteria which infect the roots and live in root nodules which fix some 1.5 million tons of nitrogen from the atmosphere into New Zealand pasture soils annually (Caradus et al., 1995). In 1990 the contribution of white clover to New Zealand's economy was estimated to be \$1.5 billion for nitrogen fixation, \$1.5 billion for its feed value, \$25 million in seed production and \$30 million for clover honey, for a total value of over \$3 billion annually (Caradus et al., 1995). No newer estimates are available.

Livestock prefer to graze white clover over grass and white clover has a higher nutritive and feeding value than grasses (Ulyatt, 1981). Lactating dairy cows have high energy needs and those that had access to 100% white clover pastures produced 1.9 to 2.0 kg milk solids per cow per day compared with 1.5 kg for cows grazing 100% grass or mixed pastures (92% grass, 8% white clover) (Cosgrove et al., 2006).

1.3 Isoflavonoids

Isoflavonoids are secondary metabolites that are derived from the phenylpropanoid pathway. The amino acid phenylalanine and malonyl CoA are essential precursors to this pathway (Winkel-Shirley, 2001). Isoflavonoids make up a large and distinctive subclass of the flavonoids with a 3-phenylchroman skeleton. There are 13 classes of isoflavonoids, briefly described below (Figure 1.2) and detailed in Dewick (1994).

Isoflavones comprise the largest group of the isoflavonoids (Dewick 1994) and are of most interest to this study since this group includes the major isoflavonoid components in white clover and red clover and have the simplest isoflavonoid structures. This group includes daidzein, genistein, formononetin and biochanin A. Red clover additionally contains the isoflavone phytoalexins pratensein, prunetin and irilone (Wu et al., 2003).

Pterocarpan isoflavonoids have an ether linkage between the 4- and 2'-positions and almost all of the pterocarpan have antifungal activity (Dewick 1994). Pterocarpan are important phytoalexins commonly found in legumes including medicarpin and maackiain and the group includes the phytoalexins phaseollin from bean and pisatin from pea (Dewick 1994).

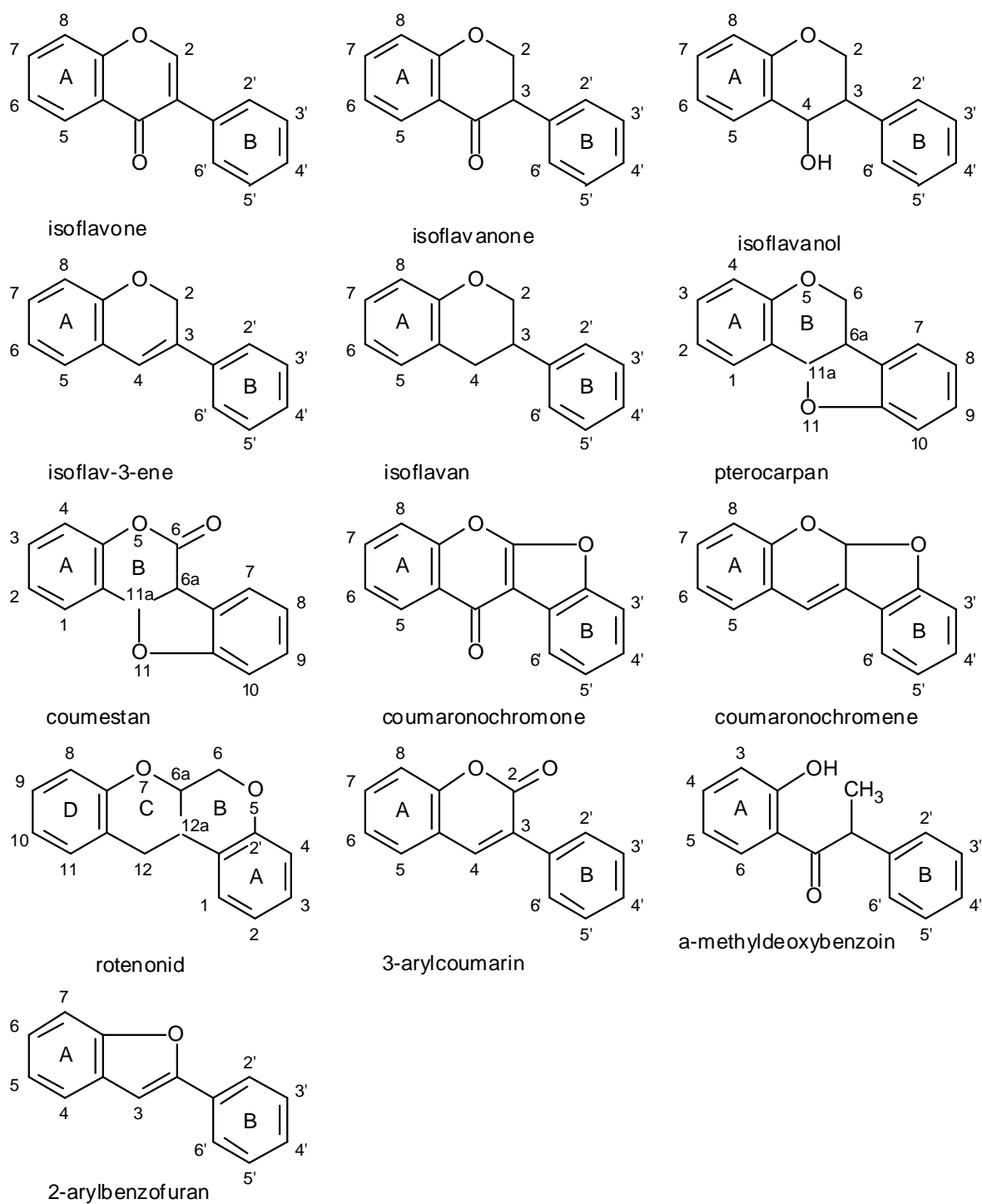


Figure 1.2 Structures of the 13 known classes of isoflavonoids.

Other classes of isoflavonoids are isoflavans, coumestans, isoflavanones, rotenoids, isoflavanols, isoflav-3-enes, 3-arylcoumarins, coumaronochromes, coumaronochromenes, α -methyldeoxybenzoin, 2-arylbenzofurans and isoflavonoid oligomers (Dewick 1994).

1.4 Plant protection and signalling

1.4.1 Isoflavonoids as plant protectors

Formononetin is a feeding deterrent for CRW and some nematodes as shown in the following studies. In a white clover population showing segregation in resistance to the stem nematode *Ditylenchus dipsaci* infection, infected meristems of resistant seedlings accumulated formononetin while susceptible seedlings did not, even though there were no significant differences in formononetin-7-O-glucoside-6'-O-malonate and medicarpin-3-O-glucoside-6'-O-malonate concentration between healthy tissues of the resistant and those of the susceptible seedlings (Cook et al., 1995). In field trials of several clover species, red clover, which had higher formononetin content than white clover, was more tolerant to damage from CRW (Cooper et al., 2003). Gerard et al. (2005) tested the interaction between CRW and one white clover ('Huia') and three red clover ('G27', 'Pawera' and 'GF131') lines with different formononetin content. The high and low formononetin red clover lines were selections from the cultivar 'Pawera' so the three lines were derived from the same genetic background. In a no-choice experiment, CRW larvae ate any of the four lines tested but gained more weight on white clover than on red clover. The same result had been reported by Byers and Kendall (1982). In another no-choice experiment, sexually mature CRW females ate any of the four lines tested and had increased mortality, laid fewer eggs and had increased abdominal fat and oil when feeding on the red clovers. When sexually immature CRW adults were given a choice of all four plants to feed on, weevils preferentially ate white clover, and the lower-formononetin red clover plants were preferred over the higher-concentration red clovers. After seven days of herbivory by CRW larvae, the formononetin content of the high-formononetin 'GF131' red clover plants more than doubled, the formononetin content of the medium-formononetin 'Pawera' increased by around 20 to 40% and the formononetin content of the low-formononetin 'G27' plants were unchanged in both leaf and root tissues (Gerard et al., 2005)

In a study assessing CRW adult survival when feeding on various red clover varieties and attraction of CRW larvae to the roots, Murray et al. (2007) found that the ‘Norseman high’ line (12,700 $\mu\text{g g}^{-1}$ dry weight (DW) formononetin) produced significantly shorter CRW adult survival times and reduced larval attraction compared with the ‘Norseman’ (6000 $\mu\text{g g}^{-1}$ DW formononetin) and ‘Norseman low’ (1200 $\mu\text{g g}^{-1}$ DW formononetin) lines.

Johnson et al. (2005) tested the response of CRW larvae to formononetin and genistein aglycones and found that they were attracted to, and moved towards, formononetin but not genistein. This suggests that the larvae might be using root-exuded formononetin as a signal to find active clover root nodules, which exuded more formononetin than inactive nodules or non-nodulated roots.

Biochanin A may also be an important feeding deterrent to CRW. When five species of clover (subterranean, white, suckling, striated and clustered clover) were tested for CRW resistance, subterranean clover was most resistant (Crush et al., 2007). Subterranean clover has moderate levels of biochanin A (about 5000 $\mu\text{g g}^{-1}$ DW) in roots and shoots but levels of formononetin lower than those active in deterring CRW in red clover (Gerard et al., 2005).

Another pest of subterranean clover is the redlegged earth mite, *Halotydeus destructor*, particularly in southern Australia. Redlegged earth mite were significantly deterred from feeding by free genistein at 0.045% fresh weight (FW) and biochanin A at 0.15% FW while the concentrations of these in leaves of the resistant subterranean clover variety ‘SE014’ are about the same at 0.05% FW for genistein and 0.17% FW for biochanin A (Wang et al., 1998). The other isoflavonoids measured also showed significant feeding deterrence – formononetin at 0.15%, formononetin glucoside at 0.5%, daidzein at 0.5%, biochanin A glucoside malonate at 0.5%, genistein glucoside malonate at 1%, 2,5,7-trihydroxy-4'-methoxyisoflavanol at 0.5% and 2,5,7,4'-tetrahydroxyisoflavanol at 0.5%, but the concentrations found in ‘SE014’ are lower than these thresholds.

Other examples of isoflavonoids acting as insect feeding deterrents include luteone and licoisoflavone A and B found in roots of *Lupinus angustifolius* (lupin) which deter feeding of two pasture scarabs *Costelytra zealandica* (grass grub) and *Heteronychus arator* (african

black beetle) as well as inhibiting sporeling growth of the fungi *Colletrichum gloeosporioides* and *Cladosporium cladosporioides* (Lane et al, 1987).

Sutherland et al. (1980) found that genistein, biochanin A and coumestrol significantly deterred *H. arator* at concentrations of 200 µg/ml. Formononetin-7-O-β-glucosyl glucoside extract from *Maackia amurensis* was found to kill brown planthopper insects via topical application or spray (Youn et al., 1991). Soybean (*Glycine max*) was resistant to the nematode *Meloidogyne incognita* due to a hypersensitive response to the nematode in the roots and production of glyceollin (Kaplan et al., 1980a) which inhibits the oxidative respiration and motility of *M. incognita in vitro* (Kaplan et al., 1980b).

Isoflavonoid biosynthesis genes are induced in soybean (Bhattacharyya and Ward, 1987) and alfalfa (*Medicago sativa*) (Paiva et al., 1994) in response to pathogenic fungi and phytoalexin isoflavonoids including sativan and medicarpin are produced, although these phytoalexin isoflavonoids are not produced in response to symbiotic mycorrhizal fungi (Paiva et al., 1994). Similarly, fungi can have isoflavonoid detoxification enzymes that impart resistance to these compounds. Isolates of *Nectria haematococca* that contained pisatin demethylase and could detoxify the phytoalexin pisatin were virulent in pea (*Pisum sativum*) while isolates lacking the gene were not (Van Etten et al., 1989).

Almost all pterocarpin isoflavonoids have antifungal activity (Dewick, 1994). Pterocarpan commonly found in legumes are medicarpin and maackiain, phaseollin, pisatin, vestitol and sativan isoflavan. Medicarpin and its metabolic pathway precursors, 2'-hydroxyformononetin and vestitone, significantly inhibited alfalfa fungal pathogens at concentrations found in elicited leaves (Blount et al., 1993). UV-B light induces the production of isoflavonoids in legumes (Hadwiger and Schwochau, 1971; Bridge and Klarman, 1973; Beggs et al., 1985) which may provide protection from oxidative damage caused by ultraviolet (UV) light by absorbing UV light and acting as an antioxidant in plants, although in white clover the flavonols quercetin and kaempferol are more strongly implicated as UV protectants (Hofmann et al. 2003).

1.4.2 Isoflavonoids as plant signals

Establishment of the legume/rhizobium symbiosis requires the plant to exude flavonoids, including isoflavonoids, from the roots (Dakora et al., 1993). The exudation and accumulation of isoflavonoids in lupin seedling was negatively correlated with nitrogen levels in the nutrient solutions that the seedlings were in (Wojtaszek et al., 1993). Genistein and isoliquiritigenin exuded by soybean are able to induce nod genes (Kape et al., 1992) and activate a protein secretion system (Süß et al., 2006). When isoflavone synthase (IFS) gene expression was silenced in soybean roots using RNA interference, the number of *Bradyrhizobium japonicum* symbioses producing root nodules decreased to 15% of the wild-type and the concentrations of genistein and daidzein in the roots were both positively correlated with the number of root nodules (Subramanian et al., 2006). These compounds are detected by rhizobial NodD proteins which activate the transcription of Nod factors (Spaink, 2000; Subramanian et al., 2007) that influence plant root hair morphology, enabling the bacteria to enter the roots (Broughton et al., 2000). Isoflavonoids are also important in the symbiotic relationship between mycorrhizal fungi and *M. sativa* (Volpin et al., 1995) and white clover (Nair et al., 1991).

1.4.3 Effects on animal health

It was first observed in Western Australia in the 1940s that ewes grazing on subterranean clover var. 'Dwalganup' had breeding problems including reduced fertility, prolapsed uterus and dystocia (difficulties giving birth). This was termed "clover disease" and hypothesised to be due to excess estrogen either ingested in, or induced by, the feed (Bennetts et al., 1946). It was not until two decades later that it was discovered that isoflavonoids, that are found in high concentrations in subterranean clover (genistein, biochanin A and formononetin combined make up 2.5 to 5.6% of DW, of which formononetin was up to 2.1% DW), were estrogenic and that this was the cause of "clover disease" (Beck, 1964). Short term grazing of estrogenic pastures from eight days before, to 17 days after, mating also reduced ewe fertility (Kelly et al., 1980). This caused a reduction in ovulation and mating and a concomitant reduction of lambing. Formononetin is demethylated and metabolised to equol by microorganisms in the animal digestive system and this estrogenic compound can build up to relatively high levels in the animal (Dixon, 2004, Setchell et al., 2002, Wang et al.,

2002). Since then, low-estrogen varieties of clover species have been bred. The cultivar 'G27' red clover (*Trifolium pratense*) was bred from the high-estrogen cultivar 'Pawera' (Rumball et al., 1997). Adams (1995) reported that high isoflavonoid *T. subterraneum* pastures reduced the fertility of ewes and cows for several weeks or months after being transferred to non-estrogenic feed. Cows on estrogenic feed had a high rate of cystic ovaries and feeding ewes estrogenic feed for prolonged periods of time could result in a permanent reduction in fertility caused by changes in the shape of the cervix. High estrogenic isoflavonoid intake has not been shown to reduce the fertility of rams, bulls or other grazing animals, nor in humans.

1.5 Isoflavonoid biosynthesis

The precursor to isoflavonoid biosynthesis, phenylalanine, is converted to cinnamic acid by phenylalanine ammonia lyase (PAL) which removes the amine group. Cinnamic acid 4-hydroxylase (C4H) adds a hydroxyl group to produce *p*-coumarate (Winkel-Shirley 2001). A coenzyme A (CoA) unit is attached at the three-carbon side chain by 4-coumarate:coenzyme A ligase (4CL) and then chalcone synthase (CHS) condenses the *p*-coumaroyl-CoA with three malonyl-CoA molecules which are circularised to make a phenyl group (Figure 1.3). The product of this reaction is naringenin-chalcone (Yu and McGonigle, 2005). Alone, CHS produces tetrahydroxychalcone and when it acts together with chalcone reductase (CHR) it produces trihydroxychalcone where the hydroxyl group at position C5 has been reduced to hydrogen (Wang, 2010). Chalcone isomerase (CHI) closes the third ring structure and produces naringenin from tetrahydroxychalcone and liquiritigenin from trihydroxychalcone. Besides being a precursor to isoflavonoids, naringenin is also a precursor for the flavonoids, condensed tannins and anthocyanins (Winkel-Shirley 2001). Both the 5-deoxy and 5-hydroxy, chalcones and isoflavonoids, are substrates of the enzymes IFS and IOMT (Figure 1.3).

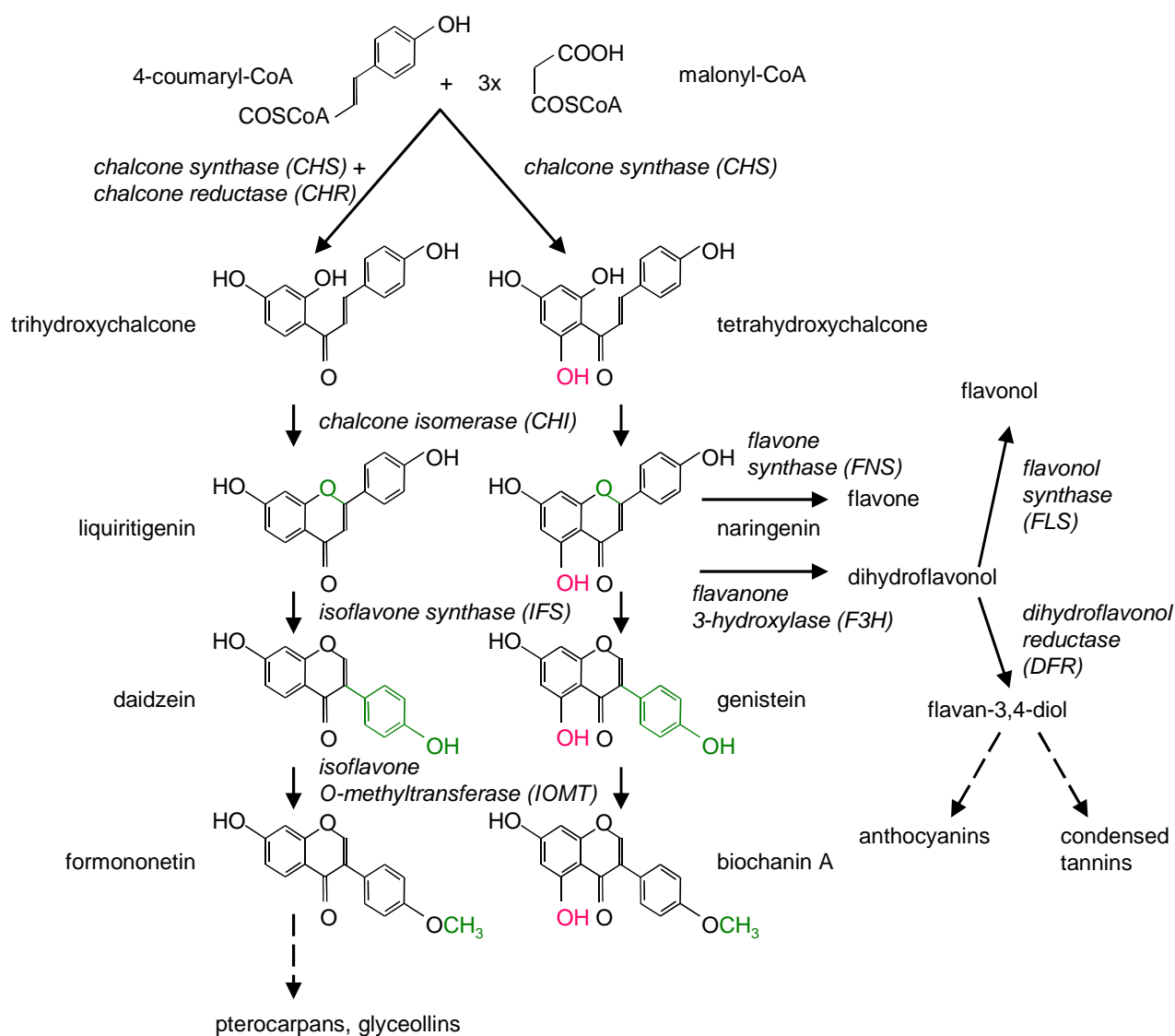


Figure 1.3 Outline of the isoflavonoid biosynthesis pathway in legumes. Figure shows the parallel processing of the 5-hydroxylated (pink) or 5-reduced chalcones and isoflavones. The key branch point of naringenin for producing other flavonoids, anthocyanins and condensed tannins is shown. Enzymes and their abbreviations are labelled in *italics*, dashed arrows indicate several enzymatic reactions that are not shown. Green features highlight the changes in the molecules.

Isoflavonoids are mostly found in glycosylated forms, in clover primarily as 7-O-glucosides and 7-O-glucoside-6''-malonates, which are conjugated by endogenous broad-substrate-range glycosyltransferases. Uridine diphosphate glycosyltransferases (UGTs) are the main enzymes that catalyse glycosylation of plant natural products (Ross et al., 2001). The glycosylated forms are less potent bioactives and more water-soluble than the aglycones, which facilitates their storage in the vacuole of the plant cell (Jones and Vogt, 2001) and may function as pools of isoflavonoids that can be released as defence compounds if required (Dakora and Philips, 1996). Indeed, early measurements of isoflavonoids were misleading because the leaf tissue was crushed at room temperature and the conjugated isoflavonoids were rapidly hydrolysed to aglycones. Placing leaves in boiling ethanol prior to grinding led to the extraction of the conjugates. This suggested that during room temperature extraction, a heat-labile hydrolytic enzyme rapidly converted the isoflavonoids into the aglycone form (Beck, 1964). This mechanism may also be active when plant tissue is being eaten by pests - isoflavonoid aglycones would be released by the maceration of the plant tissue to deter feeding. A β -glucosidase which is highly active towards isoflavone conjugates was purified and characterised from soybean (Hsieh and Graham, 2001) and lupin (Piślewska et al., 2002). Several UGTs have been identified as isoflavonoid-specific glycosyltransferases, including licorice (*Glycyrrhiza echinata*) UGT73F1 (Nagashima et al., 2004) and soybean UGT73F2 and UGT88E3 (Noguchi et al., 2007; Dhaubhadel et al., 2008).

1.5.1 Chalcone reductase and chalcone isomerase

CHR is essential for the production of 2-hydroxyisoflavonoids which are the precursors of most of the legume phytoalexins like medicarpin and formononetin. CHR is a member of the aldo-keto-reductase superfamily which catalyses NAD(P)H dependent reduction of a range of carbonyl chemicals. CHS produces chalcones in all higher plants by sequentially condensing three malonyl-CoA molecules into tetrahydroxychalcone (Ferrer et al., 1999). When it is present, CHR removes the hydroxyl group from the second malonyl-CoA, making a trihydroxy-6'-deoxychalcone. Evidence has been provided that CHS and CHR operate together as an enzyme complex, and that they are transcriptionally coactivated which supports this suggestion (Wang, 2010). Three-dimensional structures have been elucidated

for both CHS (Jez et al., 2000) and CHR (Bomati et al., 2005) which allows potential CHS-CHR interactions to be modelled.

CHR co-acts with CHS to produce 4,2',4'- trihydroxychalcone (isoliquiritigenin) from 4-coumaroyl-CoA and malonyl-CoA (Ayabe et al., 1988). This product is a precursor of the isoflavonoids daidzein and formononetin. In the absence of CHR, CHS produces 4,2',4',6'- tetrahydroxychalcone (naringenin chalcone) which is converted to naringenin by CHI (Yu and McGonigle, 2005). Naringenin is the precursor of the isoflavonoids genistein and biochanin A as well as other flavonoids. The presence or absence of CHR thus directs whether isoflavonoids are produced as 5-deoxyisoflavonoids (daidzein and formononetin) or 5-hydroxyisoflavonoids (genistein and biochanin A) (Figure 1.3).

CHI catalyses the isomerisation of chalcones into (2S)-flavanones in a stereospecific manner (Figure 1.4). Some chalcones in aqueous solution can spontaneously isomerise to the 2RS form which cannot be utilised or processed further (Jez et al., 2000). CHI has a high reaction rate which minimises the production of 2RS forms by rapidly processing chalcones (Mol et al., 1985).

Isomerisation of naringenin chalcone into naringenin occurs rapidly but converting isoliquiritigenin to liquiritigenin is slower (Jez et al., 2002) and only occurs in legumes. This activity divides the CHI enzymes into two classes: Type 1 CHIs can only isomerise naringenin chalcone to naringenin and are found in legumes and non-legumes; Type 2 CHIs can utilise isoliquiritigenin as a substrate as well and is only found in legumes. The significance of this specificity for metabolic engineering is that for any future genetic engineering experiments, a CHI enzyme that will catalyse the desired substrates needs to be chosen.

The structures of *Arabidopsis thaliana* Type 1 CHI and alfalfa Type 2 CHI have been resolved showing that the substrate bound to the active site forms a constrained configuration and is converted to the product by an acid-base catalysis mechanism (Jez and Noel, 2002; Jez et al., 2000).

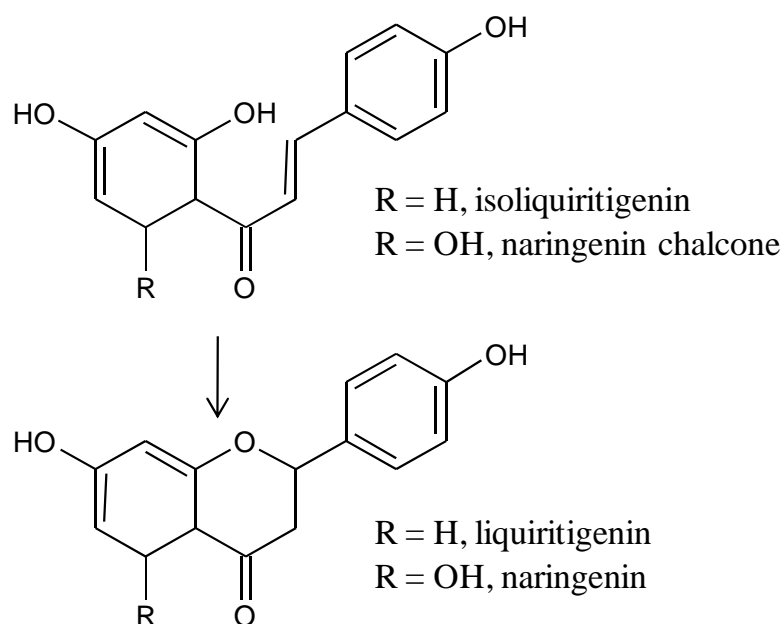


Figure 1.4 Reaction catalysed by chalcone isomerase.

Cloning of CHR would allow the creation of a vector that down-regulates this gene and the assessment of the effects on the isoflavonoid profile of transgenic clover plants carrying this vector. Given the biosynthesis pathway described in Figure 1.3, it is predicted that a down-regulation of CHR would lead to a decrease of daidzein, formononetin and any isoflavonoids where these are precursors.

The CHR protein was first isolated from soybean and the protein partially sequenced by Welle and Grisebach (1988). The corresponding CHR cDNA (P10) was first cloned by Welle et al. (1991) who also sequenced it, expressed it in *Escherichia coli* and functionally tested it. Five full length CHR sequences, which had high levels of similarity, were obtained from the legume alfalfa by Ballance and Dixon (1995) and Sallaud et al. (1995).

The genes for CHI Types 1 and 2 have been cloned from a range of plants (Blyden et al., 1991; Kimura et al., 2001; Shimada et al., 2003). CHI sequences of the same type had very similar amino acid sequences, sharing greater than 70% identity while between the two types there was about 50% identity. Shimada et al. (2003) also found that the two types were differentially expressed upon elicitor treatment.

1.5.2 Isoflavone synthase

The term IFS (isoflavone synthase) has been commonly used to describe the combined aryl migration reaction, catalyzed by 2-hydroxyisoflavanone synthase (2-HIS), and the subsequent dehydration reaction, catalyzed by the separate enzyme 2-hydroxyisoflavanone dehydratase (2-HID), which also occurs spontaneously under acidic conditions (Akashi et al. 2005). However, as most of the genes for 2-HIS are named 'IFS' and this terminology is still in common use, the name 'IFS' is used in this thesis to refer to 2-HIS.

IFS catalyses the first committed step to the production of isoflavonoids and is located in the membrane of the endoplasmic reticulum (Hagmann and Grisebach, 1984). IFS catalyses two reactions to produce isoflavonoids – a hydroxylation and an intramolecular aryl migration (Figure 1.5) (Hashim et al., 1990). The flavanone is initially converted into 2-hydroxyisoflavanone and then to an isoflavone in three steps. Firstly, IFS creates a radical at C3. Next, an intramolecular rearrangement moves the aryl group from C2 to C3. A hydroxyl group is bound to C2. Thirdly another enzyme, (2-hydroxy) isoflavanone dehydratase, converts this 2-hydroxyisoflavanone into isoflavone (Hashim et al., 1990; Sawada and Ayabe, 2005). The oxygen of the hydroxyl group bound to C2 is derived from oxygen dissolved in water and bound to the haeme group of the enzyme.

2-hydroxyisoflavanones are unstable at, and above, room temperature so they can gradually spontaneously convert into isoflavonoids (Yu and McGonigle, 2005). An isoflavone dehydratase has been isolated (Hakamatsuka et al., 1998) from kudzu (*Pueraria lobata*) suspension cells after elicitation with yeast extract which also induced CHS and IFS. The hydrophobic but soluble enzyme required no cofactors to convert 2,7,4' trihydroxy isoflavanone into 7,4' dihydroxyisoflavone (daidzein) while a control reaction with no enzyme served as a baseline to account for spontaneous dehydration of the precursor. It is likely that homologues of this enzyme exist in other plants.

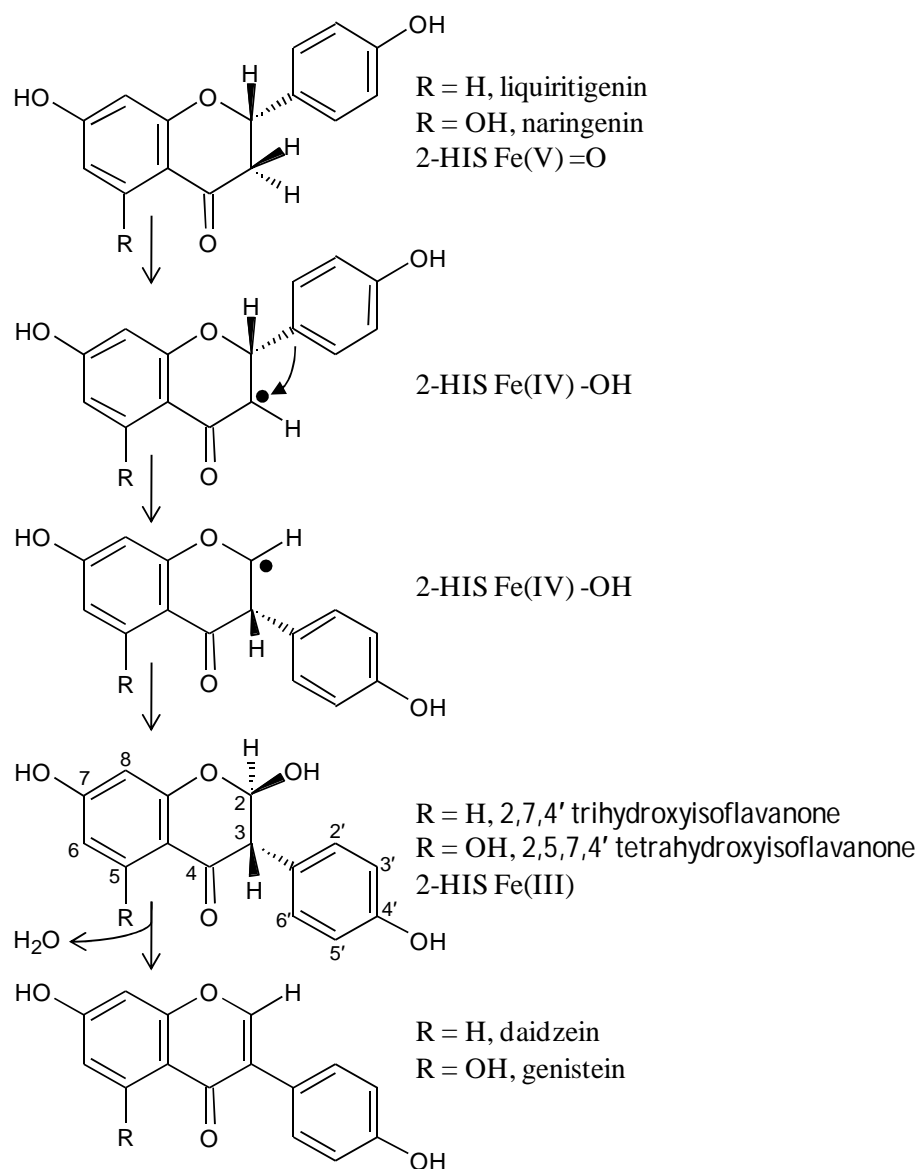


Figure 1.5 Production of genistein and daidzein by 2-hydroxy isoflavone synthase (from Hashim et al., 1990). A hydrogen is abstracted from C-3 of a flavanone substrate and, after the aryl migration, the hydroxyl is introduced at C-2. The 2-hydroxy isoflavanone is then converted into the isoflavone by 2HID.

IFS was first isolated by Hagmann and Grisebach (1984) who purified microsomes of a fungal elicitor-treated soybean cell suspension culture and determined basic properties of the enzyme. IFS was only found in the microsomal fraction, not in solution, indicating that the enzyme was bound to the endoplasmic reticulum membranes. IFS required NADPH as a

cofactor and oxygen to convert naringenin into genistein or liquiritigenin into daidzein. The requirement of oxygen and the inhibition of activity by cytochrome c indicated that it was a cytochrome P450-dependent monooxygenase (Kochs and Grisebach, 1986).

IFS genes were cloned independently by three groups. Steele et al. (1999) searched two EST libraries from soybean seeds and from *Phytophthora*-infected hypocotyl. The EST libraries were searched for sequences similar to flavone synthase II, from licorice (Akashi et al., 1998) and the product of one gene, CYP93C1v2, expressed in insect cells, was incubated with NADPH and liquiritigenin or naringenin, producing daidzein and genistein respectively. Akashi et al. (1999) cloned a full-length P450 cDNA, CYP Ge-8 (CYP93C2) from cultured licorice cells which produced the isoflavonoid-derived phytoalexin, medicarpin, on elicitation. CYP93C2 was expressed in *Saccharomyces cerevisiae* BJ2168, and the microsomal fraction was incubated with NADPH and C₁₄-labelled liquiritigenin or genistein, producing daidzein or genistein respectively. Jung et al. (2000) screened soybean EST libraries for sequences homologous to existing cytochrome p450 genes and induced by fungal infection. Two sequence variants, IFS1 and IFS2, expressed in yeast, were able to convert naringenin to genistein and liquiritigenin to daidzein as microsomal preparations. The previously sequenced soybean cytochrome P450 cyp93c1, whose function was unknown (Siminszky et al., 1999), was also assayed, converting naringenin to genistein, showing that it was a functional IFS. The soybean IFS1 was also expressed in transgenic *A. thaliana* plants which subsequently produced genistein (Jung et al., 2000).

Jung et al. (2000) went on to isolate IFS genes from other legumes by designing primers from the soybean IFS sequences. Mung bean, red clover and snow pea all produced functional IFS and with further primer design based on these sequences, IFS was isolated from white clover, hairy vetch, alfalfa and lupin. These sequences had greater than 90% nucleotide, and 95 to 99% amino acid identity. IFS was also isolated from the non-legume sugar beet which has been shown to accumulate isoflavonoids (Geigert et al., 1973). Despite the relatively distant relationship between sugar beet and soybean, the IFS genes had 95% amino acid identity which suggests a stringent requirement for the sequence of the protein able to catalyse this reaction (Jung et al., 2000).

1.5.3 Isoflavone O-methyltransferase (IOMT)

Of the over 850 aglycone isoflavonoids discovered in nature, the majority are O-methylated one or more times (Dewick, 1994). Over 80% of isoflavonoids found in alfalfa have at least one methoxy group (Bisby et al., 1994). A methoxy group is present in both of the major isoflavonoid constituents in red and white clover, biochanin A and formononetin (Dewick, 1994).

O-methyltransferases (OMTs) catalyse the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to a hydroxyl group of the substrate molecule. Joshi and Chiang (1998) distinguished two groups of small molecule OMTs in plants. Noel et al. (2003) defined Group 1 being 38 to 43 kDa proteins with no metal ion and which methylate a range of phenylpropanoids, flavonoids, coumarins and alkaloids. Group 2 are 23 to 27 kDa and require Mg^{2+} to function such as caffeoyl-CoA 3-OMT which has had its crystal structure resolved (Zubieta et al., 2001). The substrate specificity of OMTs cannot be determined from its amino acid sequence (Schroder et al., 2002) but structural models could be used to predict substrate specificity (Yang et al., 2004)

He et al. (1998) first cloned three full length IOMT cDNA clones from alfalfa. I7OMT (aka MsIOMT8) expressed in *E.coli* was a similar molecular weight as purified IOMT protein from alfalfa cell cultures and was functionally tested and capable of methylating daidzein, genistein and 6,7,4'-trihydroxyisoflavone. *In vitro*, I7OMT performed 7-O-methylation to convert daidzein into isoformononetin. *In vivo*, upon elicitation, I7OMT can also perform 4'-O-methylation of daidzein to produce formononetin, possibly because it was targeted to the endoplasmic reticulum (He and Dixon, 2000; Liu and Dixon, 2001).

Isoflavonoid 4'-O-methylation can occur on 2-hydroxyisoflavonoids, the product of IFS (Akashi et al., 2000, Deavours et al., 2006), in which case there is no production of daidzein as an intermediate. SAM: 2,7,4'-trihydroxyisoflavanone 4'-O-methyltransferase (HI4'OMT) was cloned from *G. echinata* and found to act on 2,7,4'-trihydroxyisoflavanone (and not on daidzein) to produce 2,7-dihydroxy-4'-methoxyisoflavanone, which was then dehydrated to create formononetin (Akashi et al. 2003).

Deavours et al. (2006) tested the substrate specificity of seven *Medicago truncatula* IOMTs, selected on the basis of sequence similarity of the EST to known IOMTs, and found a range and overlap of different substrate specificities, and that one or more carbon position was methylated by any one enzyme. MtIOMT3, for example, was most active methylating the C7 hydroxy group of 6,7,4'-trihydroxyisoflavone but also methylated glycitein and the 4' groups in 7,3',4'-trihydroxyisoflavone, genistein and dihydrodaidzein. MtIOMT5 acted on 2,7,4'-trihydroxyflavanone at only the 4' position and at either or both the 4' and 7 positions of dihydrodaidzein (Deavours et al., 2006). This, along with differential induction by yeast extract or methyl jasmonate, indicates that the range of IOMT enzymes performs different functions in the plant and that genetic analysis is not sufficient for determining the function of new IOMT genes at this stage - functional analysis is essential.

Examination of the protein structure of IOMTs explains some of the differences in substrate specificity where, for example, the active site of MtIOMT1 is nearly identical to that of I7OMT and they both have virtually identical activities. However, the structure of MtIOMT3 has altered hydrophobic areas and an enlarged phenolic binding pocket able to fit isoflavonoids in different configurations which explains why it is able to methylate isoflavonoids at both the 7 and 4' positions (Deavours et al., 2006).

1.6 Metabolic engineering of isoflavonoids

Traditional phenotype-based breeding is limited to the use of existing, available germplasm and it takes many years to produce a new white or red clover cultivar with a different phenotype. Marker assisted selection can speed this process up but is still limited to the available germplasm. Metabolic engineering allows the incorporation of internal genes or transgenes in engineered expression cassettes to produce a new biochemical phenotype that may not exist within any of the current genetic diversity of the species of interest. The transgenic approach can thus potentially achieve greater enhancements than breeding, and can introduce traits that cannot be introgressed by breeding.

Metabolic engineering of isoflavonoid biosynthesis has been made possible since genes for major enzymatic steps in the biochemical pathway have been cloned and sequenced. A challenge with engineering isoflavonoid metabolism is that the genetic background of model

plants is not optimised for the addition of isoflavonoid synthesis and so the yields are much lower than natural levels (Deavours and Dixon, 2005). For isoflavonoids to be produced, the precursors need to be present in the cell, so the biosynthesis of precursors must be active. In plants that do not produce isoflavonoids naturally, precursors could be produced or increased by activating the production of one of the other classes of secondary metabolites that use, for example, naringenin as a precursor, such as anthocyanins (Yu et al., 2000). If this is done in model plants with active transgenic isoflavonoid biosynthesis enzymes present, then isoflavonoids may be produced at levels comparable to those of legumes.

Tobacco (*Nicotiana tabacum*) is commonly used as a model plant for plant transformation due, in part, to its simple and efficient transformation protocol (Horsch et al., 1985). In order to use this species for functional testing of isoflavonoid biosynthesis genes, while maximising the chance of producing detectable levels of isoflavonoids, it may be useful or necessary to activate the anthocyanin pathway in the leaves. This can be achieved by expression of transcription factors that induce the production of anthocyanins or phenylpropanoids in general (Yu et al., 2000).

Jung et al. (2000) cloned the soybean IFS genes, IFS1 and IFS2, and this was the first group to introduce isoflavonoid biosynthesis into transgenic plants. Transgenic *A. thaliana* plants expressing soybean IFS1 were created and these produced genistein ($2 \mu\text{g g}^{-1}$ FW) using the endogenous naringenin precursor while the wild-type plants did not produce any isoflavonoids. Exposure to UV-B radiation induced the production of flavonoids and anthocyanins in wild-type *A. thaliana* (Caldwell et al., 1983) and exposing the transgenic *A. thaliana* plants, expressing soybean IFS1, to UV-B more than doubled the genistein concentration (Yu et al., 2000). This is an example where the production of precursors increases isoflavonoid production. Another example is in transgenic tobacco plants expressing the soybean IFS1 gene, where the flower petals, which were coloured pink by anthocyanins, produced about 130 times more genistein than leaves which do not produce any anthocyanins (Yu et al., 2000). Maize black Mexican sweet (BMS) cells expressing soybean IFS1 alone did not produce measurable quantities of genistein. CRC is a chimeric transcription factor containing maize C1 and R sequences (Bruce et al., 2000), which induces production of the entire phenylpropanoid pathway, including anthocyanins (Grotewold et al., 1998). When CRC was activated as a second transgene in the IFS1 BMS

cells, cells expressed a red phenotype and genistein was detected. Thus the level of genistein produced was proportional to, and limited by, the presence of precursors produced by an active phenylpropanoid pathway. Further, introducing a soybean CHR as well as IFS1 and CRC into maize BMS cells resulted in the production of both genistein and daidzein (Yu et al., 2000).

Liu et al. (2002) created *A. thaliana* over-expressing CHI and/or IFS and also found detectable levels of genistein in the IFS transformed plants. Increased CHI expression caused an increase in the flavonols, quercetin and kaempferol and, when IFS was also expressed, the flavonol levels dropped but the isoflavonoid levels (of around 10 nmol g⁻¹ FW) did not significantly increase.

Tian and Dixon (2006) modelled and produced an in-frame gene fusion of the soybean IFS (IFS2) and alfalfa Type II (able to use liquiritigenin as a substrate) CHI genes. IFS/CHI targeted the N-terminal IFS to the endoplasmic reticulum, while leaving the CHI free to fold to its correct conformation in the cytoplasm (Kochs and Grisebach, 1986; Bednar and Hadcock, 1988). IFS/CHI was expressed in yeast and both enzymes were functional. IFS, CHI and IFS/CHI were then transformed into separate tobacco plants and those containing IFS showed IFS enzyme activity in isolated microsomes and produced genistein in the young leaves and flower petals. While total flavonol levels rose in young leaves and did not significantly change in the petals, more genistein was produced in both tissues in five IFS/CHI plants (280 to 420 nmol g⁻¹ FW in petals - assuming DW is 10% of FW, this is approximately 76 to 114 µg g⁻¹ DW) than in five IFS plants (160 to 240 nmol g⁻¹ FW in petals - approximately 43 to 65 µg g⁻¹ DW). This is another example where the increased production of precursors, in this case by CHI, in IFS-expressing plants, increased the total production of isoflavonoids.

When IFS was introduced into an *A. thaliana* tt6/tt3 double mutant, which has structural defects in both flavanone 3-β-hydroxylase (F3H) and dihydroflavonol reductase (DFR), and produces much lower concentrations of flavonoids and anthocyanins, genistein levels were 31 to 169 nmol g⁻¹ FW which is much higher than the 5.4 nmol g⁻¹ FW produced in the non-mutant transformed plants (Liu et al., 2002). This demonstrated that a limiting factor in genistein production was competition for naringenin between IFS and F3H and when F3H

production was blocked, there was a greater amount of naringenin precursor for the IFS enzymes to convert into isoflavonoids. In this case, although flavonol biosynthesis was reduced tenfold, there was still sufficient naringenin being produced to create genistein.

Sreevidja et al. (2006) transformed rice plants with soybean IFS and were able to detect genistein in the transformed plants and IFS enzyme activity in crude homogenate protein and partially purified microsomal protein fractions from the leaves and roots. Leaf and root extracts of most transgenic rice plants were able to induce *nod* gene expression in four strains of Rhizobia, particularly *Rhizobium japonicum* USDA 110 which is maximally induced by genistein. This work is aimed at engineering a functioning rhizobia-rice symbiotic relationship to allow nitrogen fixing, leading to increased yields in this important food plant and shows that it is possible to engineer an isoflavonoid pathway in novel plants that normally do not produce any isoflavonoids.

Deavours and Dixon (2005) transformed the forage legume alfalfa, which contains endogenous IFS genes but does not produce isoflavonoids in the leaves, with *M. truncatula* MtIFS1 constitutively expressed with the 35S promoter. This resulted in the plants producing genistein, biochanin A and pratensein as well as genistin (7-O-glucosyl-genistein) and sissotrin (7-O-glucosyl-biochanin A) in the leaves. Aglycone and glycosylated genistein levels were quite low at around 13 $\mu\text{g g}^{-1}$ FW and biochanin A around 4 $\mu\text{g g}^{-1}$ FW. Stem, root and flower tissues showed no differences in flavonoid and isoflavonoid contents between MtIFS1- and vector-transformed lines, with roots producing conjugated formononetin, medicarpin and coumestrol. Stressing these alfalfa plants with 6 h of UV-B radiation, to induce the phenylpropanoid pathway, caused an increased accumulation of isoliquiritigenin, liquiritigenin and 7,4'-dihydroxyflavone in vector control line leaves, while the MtIFS1 expressing lines produced daidzein and formononetin and both lines produced detectable levels of aformosin (7-hydroxy-6,4'-methoxyisoflavone). The lack of 7,4'-dihydroxyflavone in MtIFS1 plants indicates that IFS was competing with flavone synthase for naringenin precursor molecules (Deavours and Dixon, 2005). Inoculating leaves of the MtIFS1 transgenic alfalfa with the fungal pathogen *Phoma medicaginis* caused the same changes as UV-B treatment with the addition of a significant accumulation of the antifungal phytoalexin medicarpin (Blount et al., 1993). MtIFS1 lines accumulated 4- to 7-fold higher levels of formononetin and more than 25-fold higher levels of medicarpin in

infected leaves compared with vector control lines. This indicates that metabolic engineering of isoflavonoid biosynthesis in plants for increased disease resistance is possible. Deavours and Dixon (2005) also found that the genistein-producing MtIFS1 lines did not change global gene expression in alfalfa leaves, with only one sequence from a 16086 spot microarray being significantly changed in both tested MtIFS1 lines compared with the vector control and that sequence being apparently unrelated to isoflavonoid biosynthesis.

Isoflavones are constitutively biosynthesised in soybean seeds (Graham and Graham 1991) and accumulate to high levels. Jung et al. (2003) expressed soybean IFS in soybean plants and measured the isoflavonoid content of the seed. Most plants produced isoflavonoid levels similar to those of the wild-type control and only one transformant out of fourteen produced significantly more at 1.7 times the isoflavonoid levels of the wild-type. To increase isoflavone levels further in soybean seeds, Yu et al. (2003) expressed CRC and found an approximately 2-fold increase in overall isoflavonoid levels and a greatly altered ratio of genistein: daidzein, possibly because CRC substantially increased CHR transcription. F3H was also more highly expressed and kaempferol content increased 200-fold. When CRC was expressed and F3H suppressed some seeds produced 3- to 4-fold more isoflavonoids than wild-type (and higher than any known wild-type cultivar), reduced kaempferol and returned closer to the wild-type genistein: daidzein ratio. This shows that isoflavonoid levels can be increased, even in plants and organs that naturally produce high levels of isoflavonoids, by metabolic engineering of related genes to increase precursor availability.

1.7 Project Hypotheses, Aims and Objectives

The recent introduction of clover root weevil into New Zealand highlighted the vulnerability of white clover in temperate pastures to pests and diseases which affect the productivity of the clover plants and subsequently the pastures that use the nitrogen that is fixed by white clover and the livestock who feed on them.

Tolerance or resistance to the CRW in white clover is essential for maintaining pasture productivity. Previous work suggests that isoflavonoids, in particular formononetin and biochanin A, have a feeding deterrent effect on CRW (Section 1.4.1), making high levels of isoflavonoids desirable in the plant. On the other hand, high levels of estrogenic foliar

formononetin in pastures leads to infertility in sheep so the amount, type and location of isoflavonoids in the clover plant must be balanced in order to provide both resistance to pests and maintain livestock fertility. Knowledge of the isoflavonoid biosynthesis pathways and the effects of its manipulation are essential for any transgenic approach to engineering appropriate isoflavonoid levels and also helpful to traditional or marker-assisted breeding efforts. Since little is known about the effect of overexpressing an IFS gene, the main hypotheses of this thesis were:

- IFS genes cloned from white clover are both necessary and sufficient for producing genistein in tobacco plants expressing these as transgenes.

And that:

- Increased expression of the IFS gene will increase the concentration of isoflavonoids in transgenic white clover.

The specific aims, created as milestones towards testing of these hypotheses, included:

- Identifying and quantifying isoflavonoids present in white clover populations.
- Identifying and cloning genes associated with white clover isoflavonoid biosynthesis and transferring them into a binary vector for *Agrobacterium*-based transformation into plants.
- Functional analysis of genes in model plant species and assessment of any resulting isoflavonoid profile
- Transformation of genes into white clover and assessment of resulting isoflavonoid profile

As a whole, the purpose of this research project was to explore isoflavonoid production and attempt metabolic engineering of isoflavonoids in white clover.

2 Isolation and analysis of isoflavonoid biosynthesis genes from white clover

2.1 Introduction

Isoflavonoid biosynthesis genes coding for CHR (Welle et al., 1991, Ballance and Dixon, 1995, Sallaud et al., 1995), IFS (Steele et al., 1999, Akashi et al., 1999, Jung et al., 2000) and IOMT (He et al., 1998, He and Dixon, 2000, Akashi et al. 2003, Deavours et al., 2006) have been isolated from a number of leguminous species.

For this project, cloning of IFS, IOMT and CHR from white clover was initiated with a bioinformatics approach by searching the literature for sequences of these genes from species closely related to white clover, such as from plants in the *Trifolium* genus or other legumes. These published and functionally tested sequences were used as search queries for searching the Pastoral Genomics Gene-Thresher white clover genomic DNA database for sequences analogous to these three genes. The Pastoral Genomics Gene-Thresher database is a database of genomic white clover sequences enriched for non-methylated DNA using a methodology described by Warek et al. (2005). Primers were subsequently designed to amplify and clone the genes from white clover genomic DNA.

The sequences of cloned genes were compared to estimate how many members of a gene family are present in the white clover genome, and phylogenetic analysis with other previously cloned genes allowed an assessment of which members of a gene family had been cloned and which ones were likely to exist in the genome but remain undiscovered.

High similarity of the cloned genes to previously studied genes indicated that they are the same genes and analysis, where possible, of amino acid residues that have been shown to be necessary for the functioning of the enzymes gave a greater confidence that the cloned genes were likely to be functional.

2.2 Materials and Methods

2.2.1 Locating genes in public literature, databases and the Pastoral

Genomics white clover database

Sequences for CHR, IFS and IOMT were found in the literature and through searches of public sequence databases. The public ENTREZ nucleotide database (<http://www.ncbi.nlm.nih.gov/nucleotide>) and literature databases were searched for the terms “chalcone reductase”, “IOMT” and “isoflavone synthase”.

The white clover Gene-Thresher database version PG_CLOVER, 21-04-04 was searched with BLASTN software (Altschul et al., 1997) using the following nucleotide sequences as search queries: alfalfa CHR12 (U13924) (Ballance and Dixon, 1995; Sallaud et al., 1995), white clover IFS1 (AF195814) (Jung et al., 2000) and alfalfa IOMT6 (AF000975) (He et al., 1998).

2.2.2 Contig assembly of Gene Thresher clones

Individual sequence reads (not contigs) from the PG_CLOVER Gene Thresher database with high similarity to the search queries for CHR, IFS and IOMT were aligned using Contig Express and analysed. The sequences that aligned with the fewest mismatches were assembled into a contig.

2.2.3 Structural analysis of genes

The genomic nucleotide sequences of the CHR, IFS and IOMT contigs were aligned with the coding sequence of their respective original search query (in AlignX software using the ClustalW algorithm) to determine the position of exon and intron boundaries.

2.2.4 Amplifying and cloning gene sequences from white clover DNA.

A set of primers, named ‘internal primers’, was designed within the coding sequences of the genes in the contigs to amplify and clone white clover cDNA. They did not include the very start and end sequences of the gene and thus amplified partial length genes.

A second set of primers, named ‘flanking primers’, was designed outside the coding sequences to amplify full length genomic clones of the genes. The forward flanking primer had the ‘CACC’ sequence at the 5’ end to allow it to be cloned directionally with pENTR-D-TOPO.

A third set of primers, named ‘coding sequence primers’, was designed to match the exact start and end of the coding sequences (from the ‘ATG’ start to the ‘TAA’ stop codons) to amplify and clone full length genomic sequences to be used for plant transformation experiments. This set either had Gateway *attB* sequences on the outside of the primer sequence to allow cloning into pDONR221 or ‘CACC’ on the 5’ end of the forward primer to allow the PCR product to be cloned directionally with pENTR-D-TOPO. The IOMT coding sequence primers were designed from an alignment of the 3’ sequence of IOMT contig 1 and five public IOMT gene sequences (*G. echinata* AB091685 and *M. sativa* AF000976, AF000975, AF023487 and U97125). These sequences were not perfectly conserved at the beginning and end of the gene, so degenerate primers were designed to match the consensus of the alignment with three degenerate bases in each primer, where the sequence varied between genes in the alignment.

For CHR: the internal primers were CHR_F_TrEST and CHR_R_TrEST; flanking primers were CHR_F_flanking and CHR_R_flanking; and the coding sequence primers were CHR-CDS_F and CHR-CDS_R. All primer sequences are listed in Appendix 4.

For IFS: the internal primers were IFS_F_TrEST and IFS_R_TrEST; flanking primers were IFS_F_flanking and IFS_R_flanking; and the coding sequence primers were IFS_R_mRNA1, IFS_R_mRNA2, IFS_F_mRNAd and IFS_F_mRNAd+CACC.

For IOMT: the internal primers were IOMT_F_TrEST and IOMT_R_TrEST; flanking primers were IOMT_F_flanking and IOMT_R_flanking; and coding sequence primers were IOMT_F_mRNAd, IOMT_R_mRNAd, IOMT_F_mRNAd+CACC, IOMT_F_mRNAd+attB1 and IOMT_R_mRNAd+attB.

Seven main PCR methods/experiments successfully amplified partial or full, genomic or cDNA CHR, IFS and IOMT sequences from white clover which were subsequently cloned and sequenced:

1. Standard PCR (Section 2.2.5) was used to amplify CHR, IFS and IOMT using genomic DNA from the white clover 'Sustain' genotype 6525/5 with flanking primers. Additionally, internal primers were used with cDNA extracted from three cDNA lambda libraries produced previously by Hancock (2002) from the leaf, root and stolon of white clover cultivar 'Sustain'.
2. PCR of white clover 'Sustain' genotype 6525/5 genomic DNA and 'Sustain' cDNA was performed using the high fidelity Triple Master polymerase mixture (Section 2.2.6). PCR products were purified and A-tailed for TA-TOPO cloning in pCR2.1-TOPO.
3. Partial length genomic IFS and IOMT sequences were amplified with all combinations of internal and flanking primers in a standard PCR (Section 2.2.5) with white clover 'Sustain' genotype 6525/5 genomic DNA.
4. To amplify full length genomic IFS, standard PCR (Sections 2.2.5, 2.2.1) was performed using white clover 'Sustain' genotype 6525/5 genomic DNA template and coding sequence primers IFS_F_mRNA_d and IFS_R_mRNA₂.
5. Triple Master PCR (Section 2.2.6) was performed with the IFS_F_mRNA_d+CACC and IFS_R_mRNA₂ primers and white clover 'Sustain' genotype 6525/5 genomic DNA template.
6. To clone CHR from red clover the coding sequence primers were used to amplify CHR from a genotype of the red clover variety 'Pawera' genomic DNA as well as from the white clover 'Sustain' genotype 6525/5 and a genotype of white clover 'Huia' using standard PCR (Section 2.2.5).
7. To clone the full length genomic IOMT sequence, Pfx50 PCRs (Section 2.2.7) were performed using white clover 647 genomic DNA template and degenerate coding sequence IOMT primers.

The ANT1 gene was amplified using a 50 µl Pfx50 PCR (Section 2.2.7) with tomato (*Lycopersicon esculentum* var. 'Grosse Lisse') genomic DNA and ANT1 forward and reverse primers with a 52°C annealing temperature.

2.2.5 Standard PCR protocol

Reaction mixtures (10 µl) contained 1 X PCR buffer, 0.25 U* Platinum[®] *Taq* (Invitrogen), 2.5 mM MgCl₂, 0.2 mM of each dNTP, 0.4 µM of each forward and reverse primer and 1 µl of DNA template. PCR reactions were performed in a thermocycler (MJResearch PTC200, ABI9700, ABI2720 or BioRad iCycler) running the following temperature profile: 94°C for 4 min, 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 120 s, followed by 72°C for 7 min and a 12°C hold. Extension time was increased to 1 min per kb of expected PCR product length for PCR products expected to be longer than 500 bp. Annealing temperature was set to 5 degrees below melting temperature for some primer pairs. *Unit definition: one unit of DNA polymerase incorporates 10 nmol of deoxyribonucleotide into acid-insoluble material in 30 min at 74°C.

2.2.6 Triple Master PCR

Master mix 2 (40 µl, 1 X high fidelity PCR buffer, 2.5 U Triple Master polymerase (Eppendorf) and 0.2 mM of each dNTP) was dispensed into 200 µl tubes and master mix 1 (10 µl, 0.4 µM of each forward and reverse primer, 50 ng of DNA template) pipetted on top. PCR reactions were performed in a thermocycler (MJResearch PTC200, ABI9700, ABI2720 or BioRad iCycler) running the following temperature profile: hot start, 94°C for 2 min, 30 cycles of 94°C for 20 s, 55°C for 20 s and 72°C for 90 s.

2.2.7 Pfx 50 PCR

The Pfx 50 PCR protocol used 50 µl reactions containing 1 X *Pfx50* PCR buffer, 5 U *Pfx50*[™] Polymerase (Invitrogen), additional MgCl₂ to a total concentration of 2.45 mM, 0.3 mM of each dNTP, 0.32 µM of each forward and reverse primer and 5 µl of DNA template. PCR reactions were performed in a thermocycler (MJResearch PTC200, ABI 9700, ABI 2720 or BioRad iCycler) running the following temperature profile: 94°C for 2 min, 35 cycles of 94°C for 15 s, 60°C for 30 s and 68°C for 60 s, followed by 68°C for 5 min and a 12°C hold. Extension time was increased to 1 min per kb of expected PCR product length for PCR products expected to be longer than 1000 bp. Annealing temperature

was varied for some primer pairs to 2 degrees below melting temperature but not below 60°C.

2.2.8 Agarose gel electrophoresis

Agarose (1% w/v) was dissolved and melted in hot 1 X TAE buffer (40 mM Tris-acetate, 2 mM Na₂EDTA, pH 8.5) and set in gel trays. A 20% volume of 10 X loading buffer (Appendix 1) was added to the DNA samples which were pipetted into gel wells. 10 µl of 1Kb+ ladder (Invitrogen) ready mix (Appendix 1) was loaded in the first and last wells. The samples were resolved by electrophoresis through the agarose gel at 40 to 100 V for a length of time (typically 30 min to 2 h) appropriate to the length of the gel and the resolution required. Gels were photographed using a BioRad Gel Doc or a GE Healthcare ImageQuant system and images were stored electronically.

2.2.9 Extracting DNA from agarose gels

Bands of size-separated DNA were extracted from agarose gels by excising a small plug of agarose containing the gel band and extracting the DNA using a MinElute Gel Extraction Kit (Qiagen) or by freezing the plugs in liquid nitrogen and then centrifuging at 16600 g for 5 min at room temperature and collecting the supernatant containing the DNA sample.

2.2.10 Cloning of genes

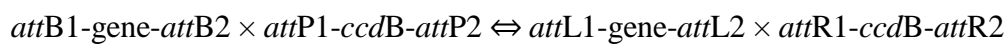
Where the forward primer had a 'CACC' sequence, PCR products were cloned into pENTR-D-TOPO vector (Invitrogen); products with *attB* sites in the primers were cloned into pDONR221 (Invitrogen) and products with no additional sequence were cloned into pCR2.1-TOPO vector (Invitrogen).

For TA-TOPO cloning of PCR products, the DNA must have a non-templated 'A' overhang at the 3' end. To ensure this is added, the DNA was incubated for 10 min at 72°C in 10 µl reactions containing 1 X PCR buffer, 0.25 U Platinum[®] *Taq* (Invitrogen), 2.5 mM MgCl₂, 0.2 mM of each dNTP and 1 to 7.6 µl of DNA.

To clone the PCR fragment, 5 ng DNA, 0.33 µl salt solution and 0.33 µl pENTR-D-TOPO or pCR2.1-TOPO vector (Invitrogen) were mixed in a 2 µl reaction and incubated for 30 min at room temperature.

2.2.11 Gateway BP and LR clonase reactions

Gateway Technology (Invitrogen) is a cloning method that uses the site-specific recombination properties of bacteriophage lambda (Landy, 1989) to move DNA sequences into multiple vector systems. The recombination sites are sequences named B, P, L and R. The Gateway Technology is schematically represented below.



(expression clone) \times (pDONR_™) \Leftrightarrow (entry clone) \times (destination vector)

The *attB* \times *attP* reaction is mediated by BP Clonase enzyme mix; the reverse *attL* \times *attR* reaction is mediated by LR Clonase enzyme mix. *ccdB* is the F plasmid-encoded gene that inhibits growth of *E. coli* (Bernard and Couturier 1992, Miki et al., 1992) and “gene” represents any DNA fragment of interest.

To move a PCR product flanked by *attB* sites into the pDONR221 vector, 5 µl reactions were performed containing 1 µl TE (Appendix 1), 1 µl BP enzyme, 1 µl (150 ng) of pDONR221 and 2 µl PCR product. The reaction was left at room temperature overnight followed by the addition of 0.5 µl of proteinase K and incubation for 10 min at 37°C.

For BP clonase reactions to move the cloned gene from pENTR-D or pDONR221 into the binary pRSh1 vector, 10 µl reactions were performed containing 2 µl BP clonase buffer, 4 µl TE, 2 µl BP enzyme, 1 µl (150 ng) of each of pDONR221 and pRSh1 vectors. The reaction was left at room temperature overnight followed by the addition of 0.5 µl of proteinase K and incubation for 10 min at 37°C.

2.2.12 Bacterial transformation

E. coli One Shot ‘DH10B’ cells or ‘TOP10’ cells (Invitrogen) were thawed on ice and 1 µl Gateway reaction products or plasmid added per tube and left for 30 min on ice. Cells were

heat-shocked for 30 s at 42°C, returned to ice for 1 min, then 250 µl SOC medium (Appendix 1) was added and incubated for 1 h at 37°C and 20 rpm rotation. 20 and 100 µl were plated out onto LB agar plates (Appendix 1) containing the appropriate antibiotic selection (100 µg ml⁻¹ ampicillin, 50 µg ml⁻¹ kanamycin or 100 µg ml⁻¹ spectinomycin) and left to grow overnight at 37°C or 2 to 3 days at room temperature.

2.2.13 Testing plasmid for insert

Bacterial colonies were picked from the LB agar plate using sterile disposable plastic pipette tips into 50 µl LB medium and resuspended. 1 µl was tested in a standard 10 µl PCR reaction using primers (M13 forward and reverse) that flanked the vector insert site. PCR reactions were run on an agarose gel and amplified bands were compared with calculated theoretical band sizes to determine whether the vector contained an appropriately sized insert.

2.2.14 Plasmid DNA extraction

10 µl of bacterial cells were inoculated into 5 ml LB medium and incubated overnight at 37°C and 200 rpm shaking. In the morning, 800 µl of bacterial culture and 200 µl glycerol were transferred to a 2 ml cryotube, mixed well and stored at -80°C. The remaining culture was centrifuged for 5 min at 3220 g to pellet the bacteria and plasmid DNA was extracted using a QIAprep Spin Miniprep Kit (Qiagen).

2.2.15 Sequencing reaction

Sequencing reactions to determine the sequence and confirm the identity of the cloned gene inserts were performed in 96-well microplates in a 20 µl volume containing a 0.875 x concentration of sequencing dilution buffer (Applied Biosystems), 1 µl BigDye v3.1 sequencing premix (Applied Biosystems), 3.2 nmol primer and 200 ng plasmid DNA template. Sequencing reactions were performed in a thermocycler (MJResearch PTC200, ABI9700, ABI2720 or BioRad iCycler) running the following temperature profile: 96°C for 1 min, 25 cycles of 96°C for 10 s, 50°C for 5 s, 60°C for 4 min, followed by a 12°C hold. M13 forward and reverse primers that flanked the insert site of cloning vectors were

generally used. If the DNA to be sequenced was longer than 1000 bp, additional sequencing reactions were performed using internal primers to obtain the full length sequence.

To remove unincorporated fluorescent dye terminators and other reaction components, 5 µl of 0.125 mM EDTA and then 62 µl of 96% v/v ethanol were added to each reaction, mixed and left at room temperature for 15 min to precipitate the sequencing products which were then collected by centrifugation for 30 min at 3220 *g* and 4°C. The microplate was subsequently inverted onto several paper towels in the centrifuge and briefly centrifuged up to 185 *g* to remove the supernatant. The precipitate was washed with 60 µl 70% v/v ethanol, and centrifuged for 15 min at 3220 *g* and 4°C. The supernatant was removed, as described above, prior to air-drying the precipitate at room temperature for 5 min. Subsequently the sequencing products were dissolved in 5 µl deionized formamide.

Sequencing reaction products were separated by capillary electrophoresis in an ABI3100 or ABI3730 sequencer (Applied Biosystems) using 36 cm or 50 cm capillary arrays and POP7 polymer.

2.2.16 Plant materials

A summary of the plant genotypes used is provided in Table 2.1 as a reference to who provided the genomic or cDNA.

Table 2.1 Summary of plant genotypes used.

Plant genotype	Species	variety	Origin of DNA	Isolated by
unknown	White clover	‘Sustain’	cDNA from root, leaf or stolon.	Hancock and Brian, (2002)
6525/5	White clover	‘Sustain’	genomic DNA	author (prior to PhD project)
647	White clover	mapping population 1	genomic DNA	author (prior to PhD project)
21125 and 20161/21	White clover	mapping population 1	cDNA	author
source of ANT1	Tomato	‘Grosse Lisse’	genomic DNA	author
red clover genotype	Red clover	‘Pawera’	genomic DNA	author
‘Huia’ genotype	White clover	‘Huia’	genomic DNA	author

2.3 Results

2.3.1 Publicly available CHR, IFS and IOMT sequences

Five full length alfalfa CHR sequences were described in Sallaud et al. (1995) and Ballance and Dixon (1995), based on sequence similarity to a partial length soybean cDNA that was functionally tested and confirmed to be a CHR (Welle et al., 1991). Searching the public ENTREZ nucleotide database for the term “chalcone reductase” resulted in 155 hits, many of which were from legumes including soybean (*Glycine max*), birdsfoot trefoil (*Lotus corniculatus*) and alfalfa (*Medicago sativa*), as well as some from *A. thaliana* and prokaryotes. Alfalfa is the closest relative of white clover amongst these organisms (Wojciechowski et al., 2004) and the alfalfa sequences from Sallaud et al. (1995) and Ballance and Dixon (1995) were among these hits.

A search of the public ENTREZ nucleotide database for the term “isoflavone synthase” found 51 hits in a range of species of bacteria and several legumes including white clover. Jung et al. (2000) cloned, sequenced and functionally verified IFS from soybean and then

cloned and sequenced IFS from a range of legumes of which the two white clover sequences were AF195814 and AF195815.

Three alfalfa IOMT sequences - AF000975 (IOMT6), U97125 (IOMT8) and AF000976 (IOMT9) - were returned in the results when searching the public ENTREZ nucleotide database for “IOMT”. IOMT8 was functionally confirmed to be active as an O-methyltransferase able to methylate genistein and daidzein (He et al. 1998).

2.3.2 CHR, IFS and IOMT from the Pastoral Genomics white clover database and contig assembly of Gene Thresher clones

The following definitions were used when assessing the quality of a database search result: The s score is a measure of the similarity of the query to the sequence shown. The E-value is a measure of the reliability of the s score. The E-value is the probability, due to chance, that there is another alignment with a similarity greater than the given s score. The typical threshold for a good E-value from a BLAST search is $e^{-5} = (10^{-5})$ or lower (<http://www.osc.edu/research/bioinformatics/FAQ/eval.e.shtml>).

Searches of the PG_GT_clover 29-07-2004 database returned these results: Using CHR U13923 (Section 2.2.1) as query, seventeen sequences with an s score of between 258 and 517 bits or an E value of between $5e^{-67}$ and e^{-145} were found. There were a further 12 sequences with an E value between e^{-5} and e^{-20} . Four of the high-score CHR hits from the PG GT database were assembled (Contig X press) with no mismatches, to make one 1966 bp contig named CHR contig 1 (Figure 2.1).

Using IFS AF195814 (Section 2.2.1) sequence as the query, several very strong hits were found. IFS Contig A assembled from the four sequences and IFS Contig B assembled from the six sequences. Contigs A and B had no mismatches in themselves but when joined to form the 3228 bp IFS Contig 3, the overlapping sequences had several mismatches (Figure 2.1).

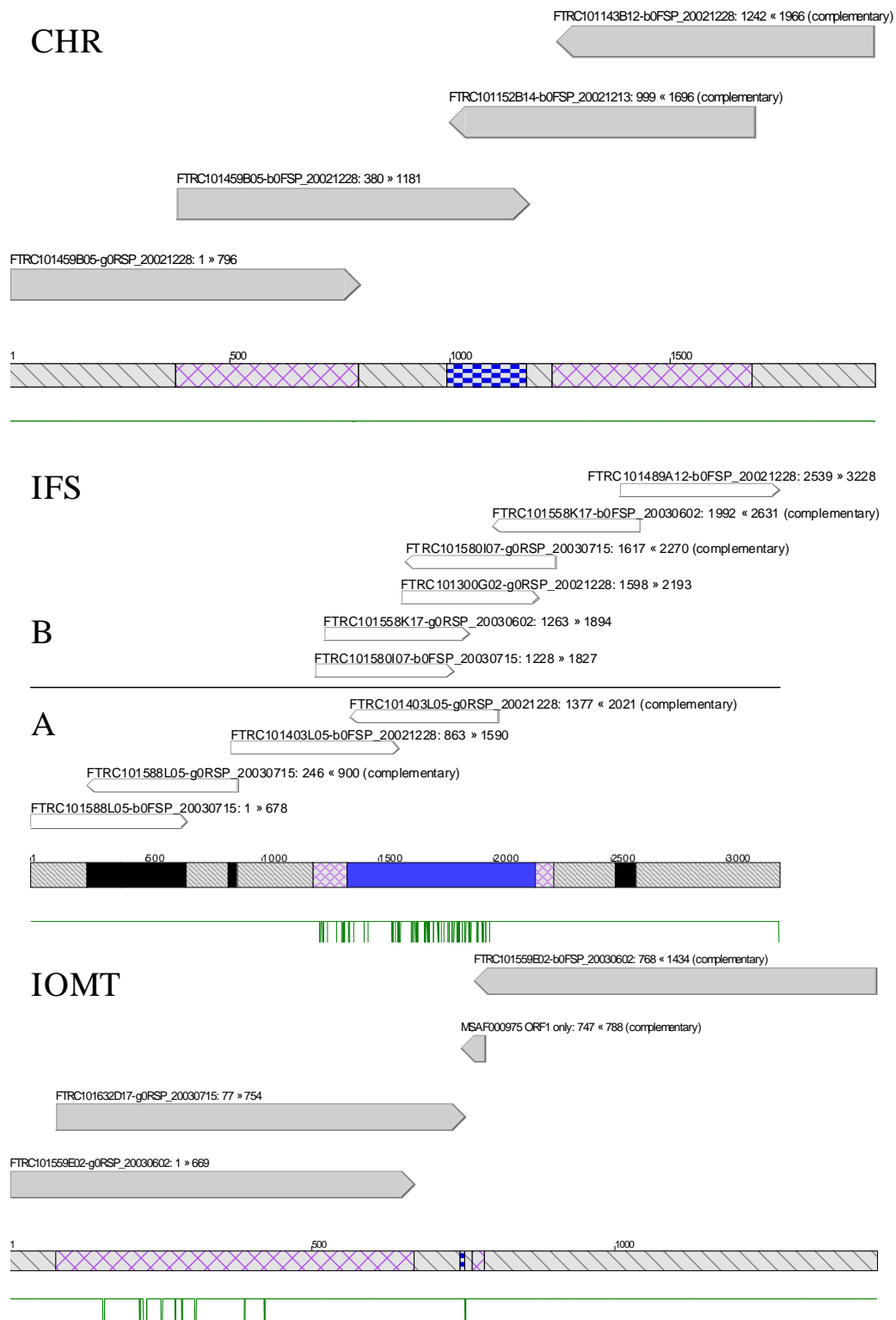


Figure 2.1 Contigs of white clover genomic sequences from a Gene-Thresher database with high similarity to CHR, IFS and IOMT search queries. Green lines indicate sequence mismatches.

Using IOMT AF000975 (Section 2.2.1) sequence as the query resulted in three strong hits which combined to cover bases 34 to 788 of the 1200 bp query as well as other lower-value hits. The three main IOMT hits assembled into a contig once a small sequence gap was bridged with 13 bp of sequence from AF000975. This contig matched most of the 5' end of the AF000975 coding sequence but not the 3' end, so the IOMT contig 1 was only a partial length IOMT sequence containing a predicted promoter and TATA signal, ORF1, intron 1 and the start of ORF2 (Figure 2.1). Compared with AF000975, the final 237 bp of open reading frame 2 was missing. IOMT Contig 1 had several single nucleotide mismatches between the two overlapping sequences and a 6 bp deletion in FTRc101632D17-g.

2.3.3 Structural analysis of genes

Aligning the nucleotide sequence of CHR contig 1 with the coding sequence of alfalfa U13924 allowed the determination of intron and exon boundaries (Figure 2.2A). Three open reading frames (ORFs) were found where the sequence of U13924 aligned with CHR contig 1 and the gaps in the alignment indicate the position of the two introns in between the ORFs. The contig contains a complete CHR gene where the ORFs have a combined length of 945 bp to produce a protein of 314 amino acids. CHR contig 1 and U13924 amino acid sequences have 94% identity, with the white clover sequence being two amino acids longer at the 3' end than the *M. sativa* sequence. Predicted promoter and TATA box features were noted from the automated annotations of the Gene Thresher sequence. A ccatttat sequence located in the predicted promoter sequence (Figure 2.2A) is present as a 2x repeat. Gene Thresher sequence FTRC101568B02-b aligned with CHR Contig 1 with some mismatches and had 100 bp more 3' sequence than CHR contig 1. This extra sequence contains a predicted promoter and TATA signal (data not shown). CHR contig 1 was submitted as the query of a BLASTN search of the nt-plant database of genbank (<http://www.ncbi.nlm.nih.gov/nuccore>) and this search returned 11 hits with an E value smaller than e^{-50} , most of which were annotated as CHR genes. The strongest hit was a *M. truncatula* clone of 113 Kb (AC141110) which had CHR in the central region and a second CHR separated by 8 kb and differing in sequence by only two base pairs.

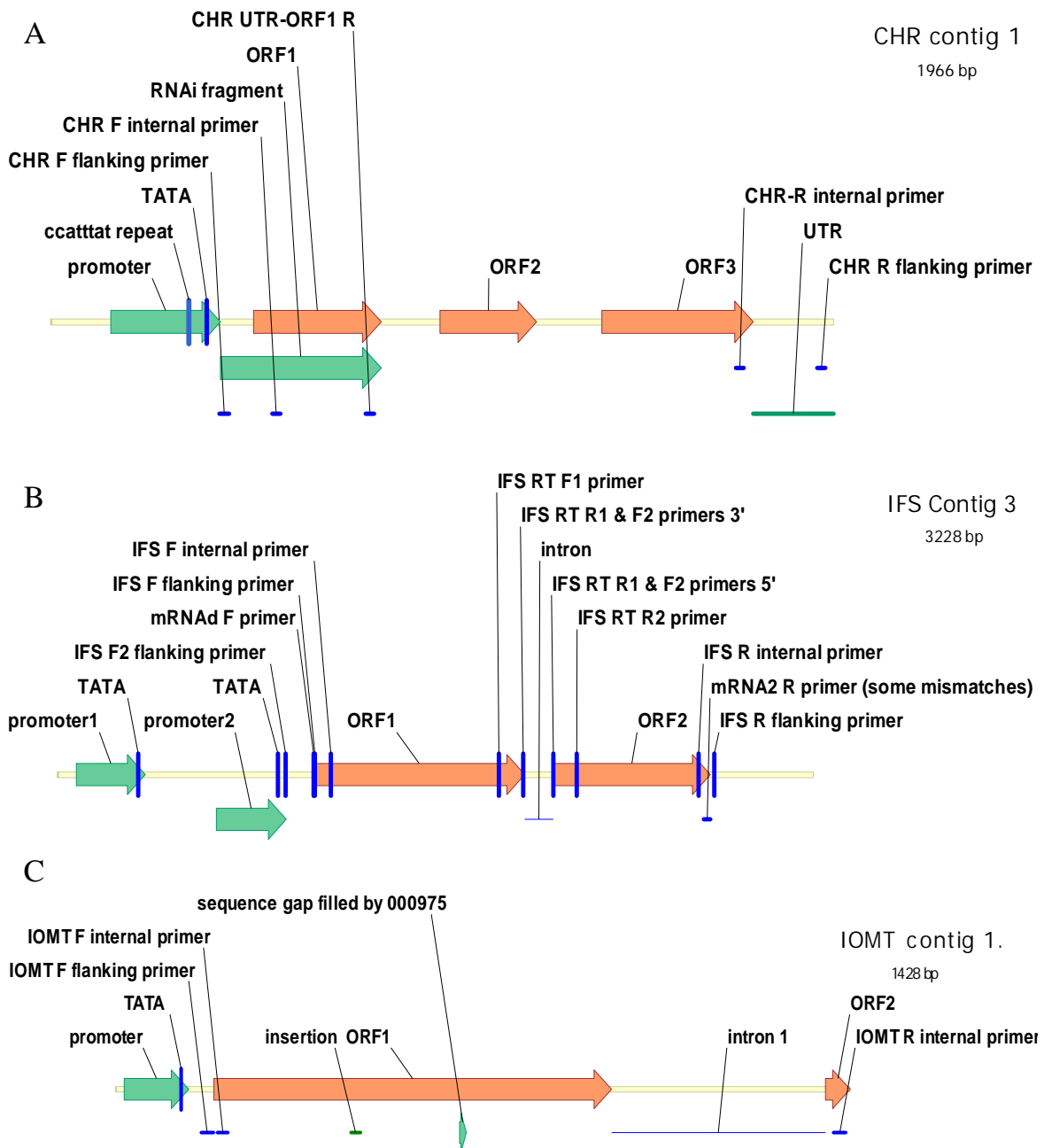


Figure 2.2 Feature maps of CHR, IFS and IOMT contigs. A, CHR contig 1. B, IFS contig 3 and C, IOMT contig 1.

IFS contig 3 contains two predicted promoters and two ORFs containing start and stop codons and so it covers a full length IFS gene. The two ORFs are 1575 bp long to form a protein 524 aa long (Figure 2.2B).

IOMT Contig 1 ORF1 is 774 bp and it contains 48 bp of ORF2. The coding sequence of AF000975 is 1059 bp long so if these coding sequences are the same total length then the contig is missing 237 bp of coding sequence (Figure 2.2C).

2.3.1 Amplifying and cloning of CHR, IFS and IOMT sequences from white clover DNA.

Sequences CHRc01 to 02 and IOMTc01 to 03 were cloned from PCR reactions following method number 1 (section 2.2.4). Cloning of IFS from these reactions was not successful. The reactions amplified predicted size products from cDNA of all three genes and gDNA of CHR, while IFS and IOMT gDNA had multiple bands, some of which were faint (Figure 2.3).

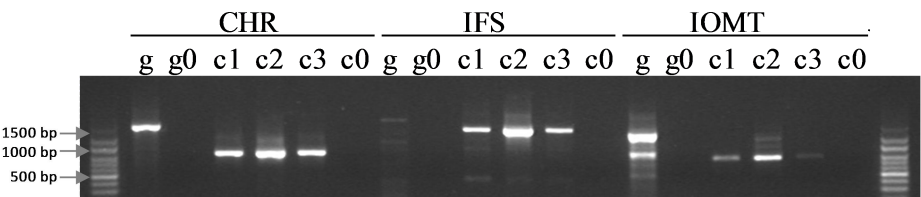


Figure 2.3 Products of PCR number 1 - CHR, IFS and IOMT from white clover ‘Sustain’ 6525/5 genomic DNA (flanking primers) and ‘Sustain’ cDNA lambda phage libraries (internal primers). g, 6525/5 genomic DNA. g0, flanking primers blank control. c1, leaf cDNA. c2, root cDNA. c3, stolon cDNA. c0, internal primers blank control

Sequences CHRc03 to 19, CHRg01 to 02 and IFSc01 to 05 were cloned from PCR reactions following method number 2 (Section 2.2.4). Both CHR PCRs amplified strongly as seen in Figure 2.4.

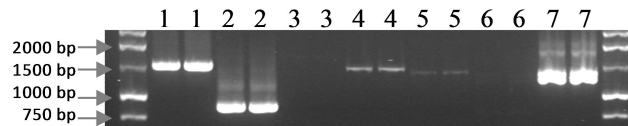


Figure 2.4 Triple Master PCR of genomic (-g) and cDNA (-c) of CHR, IFS and IOMT. 1, CHRg. 2, CHRc. 3, IFSg. 4, IFSc. 5, IOMTg. 6, IOMTc. 7, IOMTg2.

Sequences IFS 1-01 to 1-23 (n=14), IFS 3-01 to 3-24 (n=8), IOMT 5-01 to 5-23 (n=20) and IOMT 6-10 to 6-14 (n=14) were cloned from PCR reactions 1, 3, 5 and 6 following method number 3 (Section 2.2.4). These reactions produced low intensity bands (Figure 2.5) but the products cloned well.

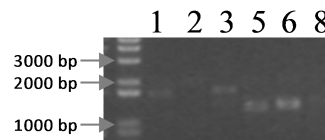


Figure 2.5 Standard PCR of genomic IFS and IOMT. 1, IFS forward internal, reverse internal primers. 2, IFS forward flanking, reverse internal primers. 3, IFS forward internal, reverse flanking primers. 5, IOMT forward internal, reverse internal primers. 6, IOMT forward flanking, reverse internal primers. 8, IOMT forward flanking, reverse flanking primers.

Amplicons from ‘Sustain’ (unknown genotype) cDNA library and ‘Sustain’ genotype 6525/5 genomic DNA (Figure 2.6) were cloned to produce cloned sequences IFS_c_M-M2 1_01 to 1_09 and full length genomic IFS_g 1_01 to 1_09 respectively, following method number 4 (Section 2.2.4).

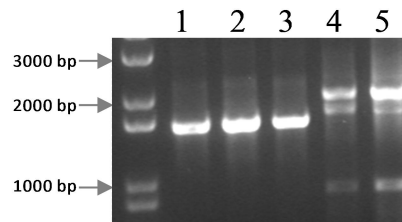


Figure 2.6 Standard PCR of IFS using IFS_F_mRNAd and IFS_R_mRNA2 primers.
1, 21125 cDNA. 2, 20161/21 cDNA. 3, 'Sustain' cDNA. 4, 6525/5 genomic DNA. 5, 647 genomic DNA.

Full length genomic IFSg 2_01 to 2_14 were cloned from PCR reactions following method number 5 (Section 2.2.4) (Figure 2.7).

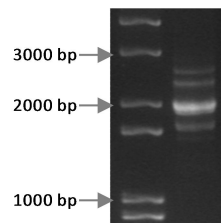


Figure 2.7 Triple Master PCR of IFS using IFS_F_mRNAd+CACC and IFS_R_mRNA2 primers with 6525/5 genomic DNA.

Sequences 6525CHRG01 to 06, 'Huia'CHRG01 and Tp1370CHRG01 were cloned from PCR reactions using white clover and red clover DNA following method number 6 (Section 2.2.4) (Figure 2.8).

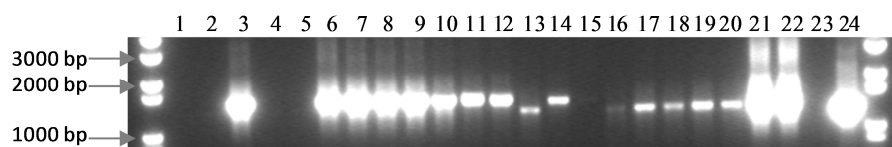


Figure 2.8 Standard PCR of CHR using CHR-CDS_GW_F and CHR-CDS_GW_R primers. This was a PCR optimisation experiment using CHR coding sequence primers. 1 to 10, 6525/5 DNA. 11 to 20, red clover ‘Pawera’ DNA. 1 to 5 and 11 to 15, annealing temperatures of 46°C, 48.7°C, 52°C (on a different thermocycler), 54.7°C, 60°C. 6 and 16, 3.5mM Mg. 7 to 8 and 17 to 18, 2% v/v and 5% v/v DMSO added, 9 to 10 and 19 to 20, 2% v/v and 5% v/v formamide added. 21, other 6525/5 DNA sample. 22, other red clover DNA sample (‘rep13’). 23 no-DNA blank control. 24, positive control.

Sequences IOMT 647 01 to 32 were cloned from PCR reactions following method number 7 (Section 2.2.4) (Figure 2.9).

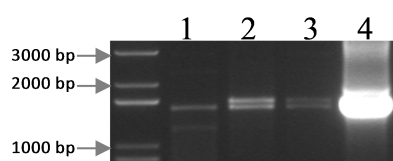


Figure 2.9 Pfx50 PCR of IOMT using IOMT_F_mRNAd+attB1 and IOMT_R_mRNAd+attB2 primers and genomic DNA. 1, 6525/5. 2, 647. 3, P06A. 4, positive control

Table 2.2 A summary of CHR, IFS and IOMT genes cloned.

Gene names and numbers	Full length gene	Predicted length (bp)	Actual lengths (bp)	Vector cloned into	DNA source	PCR method number	Polymerase used in PCR	Primers used in PCR
CHRC 01 to 02	no	874	874, 634	pENTR/D-TOPO	'Sustain' root cDNA	1	Platinum Taq	internal
CHRC 03 to 19	no	874	874	pCR2.1-TOPO	'Sustain' root cDNA	2	Triple Master	internal
CHRG 01 to 02	yes	1524	1565, 1522	pENTR-D-TOPO	'Sustain' 6525/5 genomic	2	Triple Master	flanking
6525CHRG 01 to 06	yes	1524	1426, 1430	pDONR 221	'Sustain' 6525/5 genomic	6	Platinum Taq	coding sequence
IFSc 01 to 05	no	1471	1471	pCR2.1-TOPO	'Sustain' root cDNA	2	Triple Master	internal
IFSc_M-M2 1_01 to 1_09	yes	1575	1566 to 1572	pCR2.1-TOPO	'Sustain' root cDNA	4	Platinum Taq	coding sequence
IFS 1-01 to 1-23 (n=14)	no	1593	1304, 1582, 1585, 1590	pCR2.1-TOPO	'Sustain' 6525/5 genomic	3	Platinum Taq	internal
IFS 3-01 to 3-24 (n=8)	no	1660	1660	pCR2.1-TOPO	'Sustain' 6525/5 genomic	3	Platinum Taq	F internal R flanking
IFSg M-M2 1-01 to 1-09	yes	1697	1688 to 1698	pCR2.1-TOPO	'Sustain' 6525/5 genomic	4	Platinum Taq	coding sequence
IFSg M-M2 2-01 to 2-14	yes	1697	1572 to 2006	pENTR-D-TOPO	'Sustain' 6525/5 genomic	5	Triple Master	coding sequence
IOMTc 01 to 03	no	807	801, 806	pENTR-D-TOPO	'Sustain' root cDNA	1	Platinum Taq	internal
IOMT 5-01 to 5-23 (n=20)	no	1223	1233	pCR2.1-TOPO	'Sustain' 6525/5 genomic	3	Platinum Taq	internal
IOMT 6-10 to 6-14 (n=14)	no	1255	1268, 1400	pCR2.1-TOPO	'Sustain' 6525/5 genomic	3	Platinum Taq	F flanking R internal
IOMTg 01 to 32 (n=32)	yes	1491	1491, 1497, 1611	pDONR 221	white clover 647 genomic	7	Pfx50	coding sequence

Of the genes that were cloned into pCR2.1, pENTR-D and pDONR221 (Table 2.2), the following genes were transferred into the binary vector pRSh1, suitable for transformation into plants via *A. tumefaciens*: ANT1, CHRG02, IFS/CHI number 01 to 06, IFSg 2_03, IFSg 2_10, IFSg 2_11, IFSg 2_12, IOMTg10, IOMTg12, IOMTg22.

2.3.2 Analysis of cloned genes.

When comparing cloned sequences of the same gene, four classes of sequence variations were encountered. ① A putative SNP was when two different nucleotide bases occurred in the same position in two or more sequences. This type of variation was most likely a real single nucleotide polymorphism (SNP) that was used to distinguish between different sequence variants or sequence variants of the same gene. ② A PCR mutation was when a single nucleotide base varied from all of the other sequences and it was present in only one sequence. This type of sequence variation is believed to be a mutation caused by the imperfect fidelity of the polymerase when it was copying that clone and this class of sequence variation varies in number according to the fidelity of the DNA polymerase - from about one per 700 bp with Platinum Taq and a similar number with Triple Master, to one in 43000 bp with Pfx50. ③ A primer mutation was a variation in the first few base pairs of a cloned sequence using coding sequence primers, caused by imperfectly synthesised PCR primers. ④ Indels are insertions or deletions when comparing two or more sequences. They were common in introns and also occurred in exons such as in IOMT where a 6 bp sequence deletion was present in one sequence variant and absent in another (Section 2.3.5). The length of these indels was usually in multiples of three nucleotides so that the subsequent translated amino acid sequence of the polypeptide remained in frame. Indels also distinguished different sequence variants of a gene.

2.3.3 Chalcone reductase

The sequence of CHRg02 matches CHR Contig 1 along its entire length with only a single base pair difference. CHRg02 has a coding sequence of 945 bp/ 314 amino acids with 312 bp of sequence in two introns. CHRg02 and alfalfa U13924 are very similar with 94% of their amino acid sequences being identical. 6525CHRg03 and 6525CHRg05 are 97% identical and vary from CHRg01 and 02 (which are 95% identical) only by SNPs and intron length.

In 18 CHR cDNA clones of 874 bp length, 24 putative SNPs occurred in two or more of the sequences and are believed to mark different sequence variants of CHR. All putative SNPs were spatially separated and six of the putative SNPs caused a change in the amino acid

sequence. These putative SNPs separated the sequences into five different sequence variants from one plant genotype (Figure 2.10). Sequencing of genomic clones from a different genotype (6525/5) of the cultivar ‘Sustain’ resulted in a further four sequence variants which did not match the cDNA clones exactly although CHRg02 and CHRc07 are very similar with only four putative SNPs between them.

Base pair number	37	61	124	130	178	190	194	196	241	274	398	496	508	550	607	616	619	643	725	766	775	799	821	850
CHRC16	T	C	T	T	C	G	G	A	C	A	G	C	G	G	G	A	T	G	A	T	G	T	A	C
CHRC11	T	C	T	T	C	G	G	A	C	A	G	C	G	G	G	A	T	G	A	T	G	T	A	C
CHRC19	T	C	T	T	C	G	G	A	C	A	G	C	G	G	G	A	T	G	A	T	G	T	A	C
CHRC09	T	C	T	T	C	G	G	A	C	A	G	C	G	G	G	A	T	G	A	T	G	T	A	C
CHRC05	T	C	T	T	C	G	G	A	C	A	G	C	G	G	G	A	T	G	A	T	G	T	A	C
CHRC17r	A	C	T	C	T	A	C	A	C	G	C	C	G	G	A	A	C	A	G	T	G	A	G	T
CHRC15	A	C	T	C	T	A	C	A	C	G	C	C	G	G	A	A	C	A	G	T	G	A	G	T
CHRC14r	A	C	T	C	T	A	C	A	C	G	C	C	G	G	A	A	C	A	G	T	G	A	G	T
CHRC08	A	C	T	C	T	A	C	A	C	G	C	C	G	G	A	A	C	A	G	T	G	A	G	T
CHRC01	A	C	T	C	T	A	C	A	C	G	C	C	G	G	A	A	C	A	G	T	G	A	G	T
CHRC18	A	C	A	C	C	A	G	G	C	G	G	T	T	G	A	A	C	A	G	T	G	A	G	T
CHRC02	A	C	A	C	C	A	G	G	C	G	G	T	T	G	A	A	C							
CHRC10r	A	C	T	C	C	A	G	G	C	G	G	T	T	G	A	A	C	A	G	T	G	A	G	T
CHRC13	A	C	T	C	C	A	G	G	C	G	G	T	T	G	A	A	C	A	G	T	G	A	G	T
CHRC04r	A	C	T	C	C	A	G	G	C	G	G	T	T	G	A	A	C	A	G	T	G	A	G	T
CHRC12r	T	T	T	C	C	A	G	G	T	G	G	C	G	A	G	C	C	G	G	C	A	T	G	T
CHRC06r	T	T	T	C	C	A	G	G	T	G	G	C	G	A	G	C	C	G	G	C	A	T	G	T
CHRC07	T	T	T	C	T	A	G	C	C	G	C	C	A	G	G	A	C	G	G	C	G	T	A	C
AA difference?	n	n	n	n	n	n	y	y	y	n	y	n	n	n	n	n	n	n	y	n	n	n	y	n

Figure 2.10 Illustration of putative SNPs in CHRC clones from one unknown ‘Sustain’ genotype that allow the grouping of the sequences into five variants.

Alignment of all genomic and cDNA CHR clones separated the sequences into nine sequence variants, five sequence variants from a root cDNA library of an unknown genotype of the cultivar ‘Sustain’ (Hancock and Brian, 2002) and four sequence variants from genomic DNA (from ‘Sustain’ genotype 6525/5). These were arranged in a phylogenetic tree by Geneious software (Drummond et al., 2010) to show the similarity between different sequence variants (Figure 2.11).

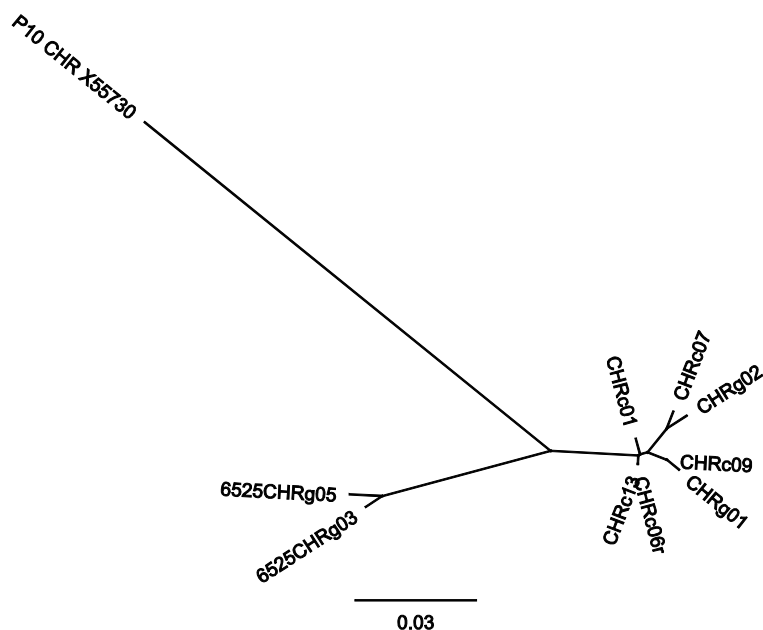


Figure 2.11 Phylogenetic tree of the amino acid sequence of representative sequences from nine CHR sequence variants from white clover and the functionally confirmed CHR P10 from soybean.

Analysis of the crystal structure of *M. sativa* CHR12 revealed the active site of CHR to be the ‘catalytic tetrad’ of Asp 53, Tyr 58, Lys 87, and His 120 (Bomati et al. 2005). An amino acid sequence alignment of representatives from the nine white clover CHR sequence variants with *M. sativa* CHR12 (Ballance and Dixon, 1995) shows that the ‘catalytic tetrad’ is present in all of the white clover sequence variants (Figure 2.12).

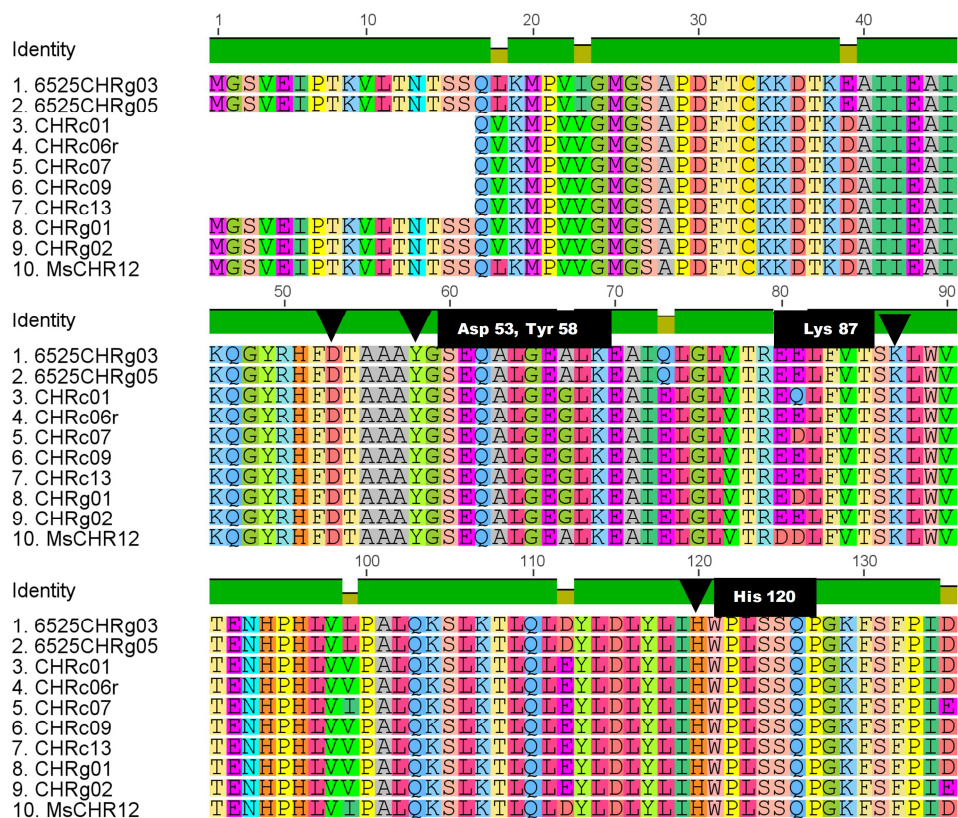


Figure 2.12 Alignment of amino acid sequence of *M. sativa* CHR12 and representative sequences from nine CHR sequence variants from white clover, showing amino acid residues of the ‘catalytic tetrad’ that are essential for CHR functioning.

2.3.4 Isoflavone synthase

IFS Contig 3 had a coding sequence of 1575 bp and 574 amino acids in two exons spanning a 122 bp intron (Figure 2.2). IFS2_12 has a longer 368 bp intron (Appendix 3 lists the full sequence of IFS2_12). The sequences that align most closely to IFS2_12 were derived from its close relatives, red clover AY253284, barrel medic AY167424 and pea AF532999 which had 92.8%, 91.3% and 88.6% identical amino acid sequence, respectively. Six of the sequence variants were cloned from the ‘Sustain’ root cDNA library and the other seven variants were cloned from the ‘Sustain’ 6525/5 genotype genomic DNA (Table 2.2), so there are at least seven IFS sequence variants in that one white clover genotype.

Comparison with 20 IFS sequences cloned from a range of different plant species (Jung et al., 2000) show that the white clover IFS genes are in two groups yet still very closely related to each other and quite distinct from the genes cloned by Jung, which also cluster closely together, in two groups (Figure 2.14).

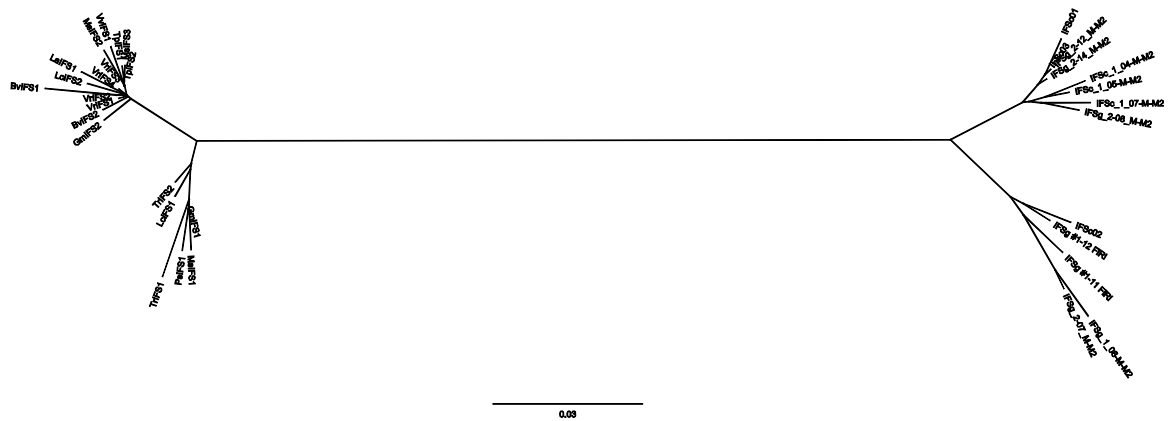


Figure 2.14 Phylogenetic tree of the amino acid sequence of representative sequences from 13 IFS sequence variants from white clover and 20 IFS genes from Jung et al. (2000) Bv, *Beta vulgaris*. GM, *G. max*. La, *Lupinus albus*. Lc, *Lens culinaris*. Ms, *M. sativa*. Mt, *M. truncatula*. Ps, *Pisum sativum*. Tp, *Trifolium pratense*. Vr, *Vigna radiata*. Vv, *Vicia villosa*. Other genes are from *T. repens*.

Alignment of the 13 white clover sequence variants with *Glycyrrhiza echinata* CYP93C2 shows that three amino acid residues that are essential for the correct functioning of the IFS gene, Ser 310, Lys 375 (Sawada et al., 2002) and Leu 371 (Sawada and Ayabe, 2005) were present in all of the white clover sequence variants (Figure 2.15).

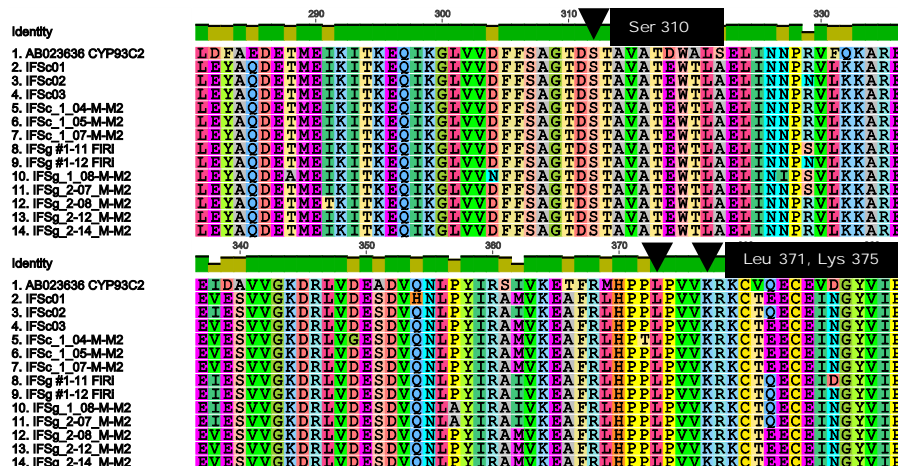


Figure 2.15 Alignment of *G. echinata* CYP93C2 and amino acid sequence of representative sequences from 13 putative IFS sequence variants from white clover showing amino acid residues that are essential for IFS functioning.

2.3.5 Isoflavone O-methyltransferase

Out of 29 proofread and manually checked IOMT sequences, there were three sequence variants in the genomic DNA of white clover genotype 647. These have a number of single nucleotide polymorphisms (SNPs) but are overall very similar with 93% consensus, 89% identity. The three sequence variants vary in length (1491, 1497 and 1611 bp) due to indels in the single intron.

When translated, the three sequence variants show 98.3% identity between the 358 amino acids. The 1611 bp sequence variant (IOMTg10) has only five amino acid changes from the 1491 bp sequence variant (IOMTg12) and the 1497 bp sequence variant (IOMTg22) only is one amino acid different to the 1491 bp sequence variant (which is the consensus sequence of the three). The 1491 bp sequence variant is only one amino acid different to the first ORF of the IOMT gene in the Gene Thresher database and is also identical in length to the IOMT contig that was assembled from the Gene Thresher database.

Sequencing of genomic clones from white clover ‘Sustain’ genotype 6525/5 and cDNA from an unknown ‘Sustain’ genotype increases the number of sequence variants to eight.

Compared with public IOMT sequences, the IOMT clones isolated are distinct but very similar (Figure 2.16). For example, IOMTg12 is 87% identical to alfalfa IOMT6 in amino

acid sequence. The changes consist of several single amino acid changes and an insertion of five amino acids in IOMTg12.

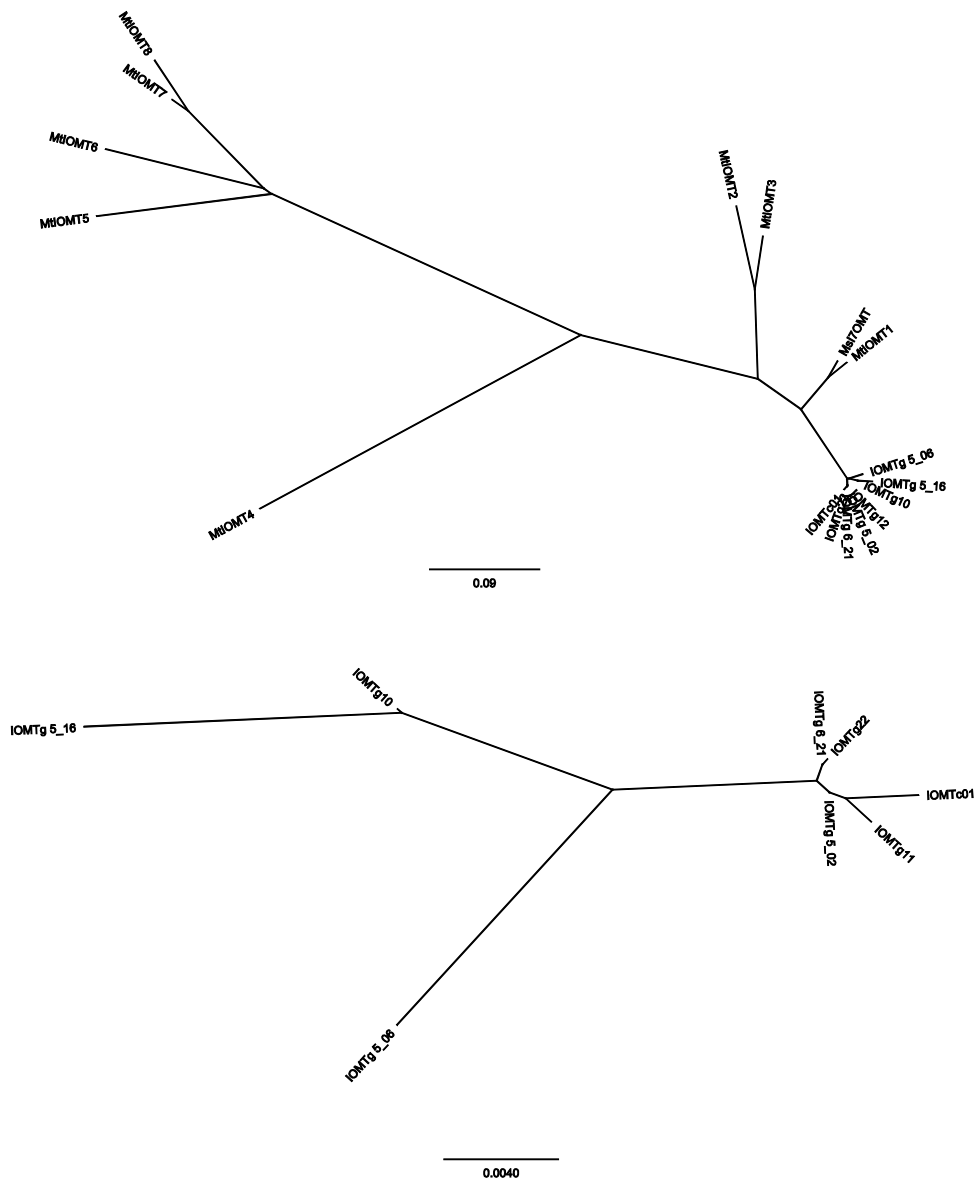


Figure 2.16 Phylogenetic tree of the amino acid sequence of representative sequences from eight IOMT sequence variants from white clover and nine functionally confirmed public IOMT genes. Upper, all genes. Lower, white clover genes only. Mt, *M. truncatula*, Ms, *M. sativa*, other genes are from *T. repens*

Alignment of white clover sequence variants IOMTg10, 12 and 22 with MsI7OMT shows that four amino acid residues that shape the enzymatic reaction pocket, Tyr 25, Tyr 127, Asn 310 and Met 311 (Deavours et al., 2006) and three amino acid residues that are involved in SAM-binding Lys 253, Trp 259 and catalysis, His 257 (Zubieta et al., 2001) were present in all of these sequence variants (Figure 2.17) (residue numbering is of MsI7OMT).

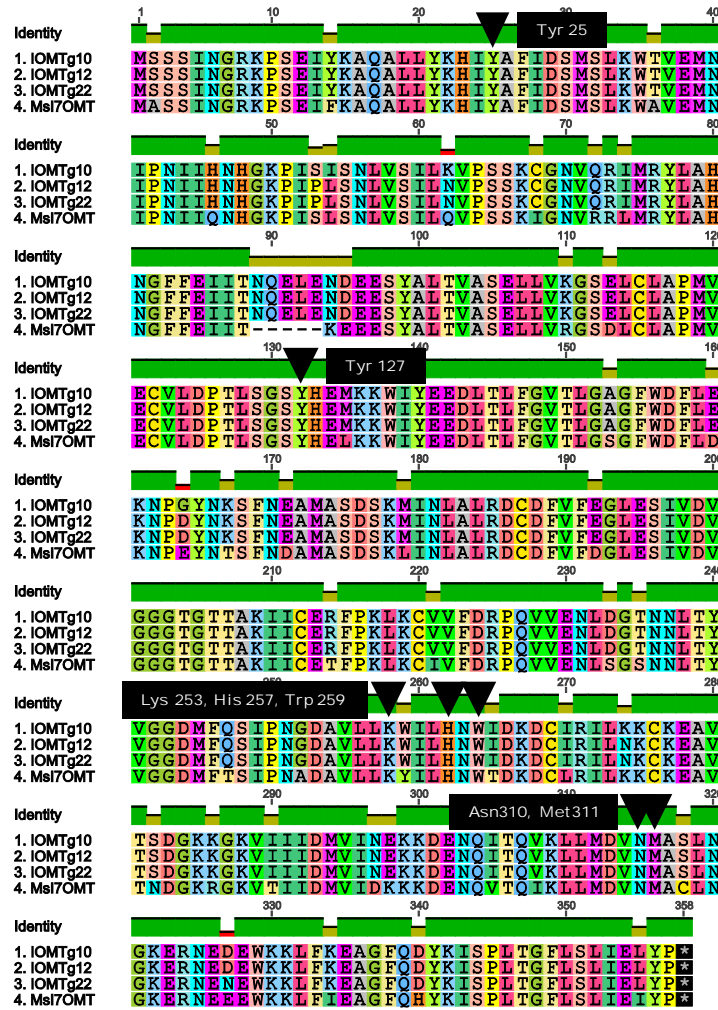


Figure 2.17 Alignment of MsI7OMT and IOMTg10, 12 and 22 showing amino acid residues that shape the enzymatic reaction pocket (Deavours et al., 2006) and are involved in SAM-binding and catalysis (Zubieta et al., 2001).

2.3.6 Cloning of ANT1 gene

Anthocyanin mutant 1 (ANT1) was isolated by Mathews et al. (2003) as a gene containing two introns and 825 bp of coding sequence cloned from the genomic DNA of the tomato *L. esculentum* var. 'Micro-Tom'. Coding sequence primers containing flanking attB sequences were designed to the start and end of the coding sequence and used to amplify the full length ANT1 gene from tomato *L. esculentum* var. 'Grosse Lisse' DNA (Section 2.2.4). A single band of a Pfx50 PCR product of approximately 1000 bp (Figure 2.18) was cloned into pDONR221 and sixteen of the clones sequenced. All clones had identical 1012 bp inserts and the coding sequence of the insert was identical to the ANT1 (AY348870) sequence published by Mathews et al. (2003).

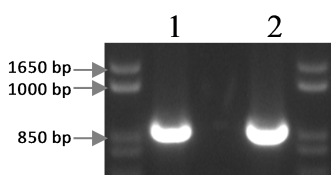


Figure 2.18 PCR of ANT1 using 1, Pfx polymerase and 2, Platinum Taq.

2.3.7 Isoflavone synthase / chalcone isomerase

The translational fusion gene IFS/CHI (soybean isoflavone synthase / alfalfa chalcone isomerase) was received from the Noble Foundation (Ardmore, Oklahoma, USA) where it had been developed by Tian and Dixon (2006). At the time the gene was received, the cloned white clover IFS genes had not yet been shown to be functional. The IFS/CHI fusion gene was therefore used as a positive control as it has been shown to be functional by Tian and Dixon (2006). IFS/CHI was received in the binary vector pBI121 and was amplified by PCR and transferred into Gateway vectors pDONR221 and pRSh1.

2.4 Discussion

2.4.1 CHR

CHR sequences from alfalfa had previously been published by Sallaud et al. (1995) and Ballance and Dixon (1995) and the expressed proteins were shown to have CHR activity.

Alfalfa is one of the most closely related legumes to white clover (Wojciechowski et al., 2004). Alfalfa and white clover genes are thus likely to have relatively high sequence similarity which makes alfalfa genes good queries for searching the white clover Gene-Thresher database for homologous white clover genes. To design primers for PCR-amplification of the white clover genes, the existing white clover DNA database was searched to identify white clover CHR sequences. White clover is a genetically diverse species so the gene sequences in the Gene-Thresher database, which were derived from an unknown genotype in the USA, are expected to vary from the genes in the plants in this study. However, the white clover database sequences were likely to be closer in sequence to the genes targeted for cloning than the alfalfa genes and thus primers designed to these sequences were more likely to amplify the desired genes. Phylogenetic analysis confirmed that the Gene-Thresher sequences were more closely related to the cloned genes than the alfalfa sequences (Figure 2.11).

The five published alfalfa CHR sequences MsCHRIa (X82366), MsCHRIc (X82367), and MsCHR2a (X82368) (Sallaud et al., 1995) and CHR12 (U13924) and CHR7 (U13925) (Ballance and Dixon, 1995) had high levels of nucleotide sequence similarity - over 98% consensus and over 90% identity in the overlapping areas - which meant that any one of them would be an equally good sequence to search the white clover genome database. The Gene Thresher database contained sequences from several different sequence variants of any of the isoflavonoid biosynthesis genes, so if genomic sequences that did not match perfectly in the overlaps were assembled, this would result in a chimeric contig where the sequences from different but similar sequence variants were joined together. This was the case for IFS and IOMT but not for CHR, for which a single contig was perfectly assembled (Figure 2.1).

Flanking sequences are non-coding and generally have little or no selective pressure to maintain a conserved sequence or function (Lockton and Gaut, 2005). They can thus vary more than the coding sequence between different sequence variants created by duplication and the same sequence variant between different genotypes of white clover and are less likely to be good sites for primers designed to amplify the gene. CHR was an exception where one complete full length sequence variant was assembled from four Gene Thresher database sequences (CHR contig 1) and the flanking sequence of this gene was sufficiently similar to that of the white clover genomic DNA template, that the designed flanking PCR

primers could anneal to the genomic DNA sequence and successfully amplify a full length white clover CHR gene. The three full length sequenced CHR clones matched the sequence of the CHR contig 1 along the entire length with only one nucleotide difference. Comparison with the coding sequence of *M. sativa* CHR12 showed three open reading frames surrounding two introns. The CHR contig 1 contains a predicted promoter, TATA box, ATG start codon and TAA stop codon. These comparisons with existing, functionally tested, CHR genes are evidence suggesting that it is a full length, functional gene (Figure 2.2).

Further evidence that the cloned genes are functional came from alignment of the nine sequence variants with *M. sativa* CHR12 which showed that all four of the amino acids from the 'catalytic tetrad' were present in all of the white clover sequences (Figure 2.12). A 'ccatttat' sequence located in the predicted promoter sequence (Figure 2.2) was present as a 2x repeat and this may affect the function of the promoter. DNA minisatellites present in regulatory elements may affect the expression of the associated gene (Meagher and Vassiliadis, 2005). This could be an interesting subject to follow up in future work, by transforming white clover with a transgenic marker gene, like GUS, preceded by this regulatory region as well as the same construct with only a single 'ccatttat' sequence. The two expression patterns would indicate the regulatory effect of this minisatellite.

A *M. truncatula* clone of 113 kb (AC141110) contains a CHR gene in the central region which makes it possible to determine what other genes may be clustered around this CHR. There is a second CHR gene present in this sequence, separated from the first by 8 kb and differing in sequence by only two base pairs, suggesting that this pair was created by gene duplication which can then evolve separately in the future. It is possible that such ordering or clustering is present in other legumes including white clover. Gene Thresher sequence FTTC101568B02-b, which aligned with CHR Contig 1 with some mismatches, had 100 bp more 3' sequence than CHR contig 1. This extra sequence contains a predicted promoter and TATA signal which probably represent the beginning of a new gene, perhaps CHR, in the 3' direction from at least one of the CHR genes. Sequencing of further regions adjacent to CHR in white clover to see if this is the case was not attempted because it fell outside the scope of this work, and existing Gene Thresher contigs are not long enough to detect this through a bioinformatics approach.

Nine different CHR sequence variants were cloned and sequenced from white clover based on single nucleotide polymorphisms in the coding sequence, although CHRg02 and CHRc07 vary by only four SNPs and may be the same sequence variant. The sequence variants were cloned from three different genotypes (20161/21, 6525/5, and an unknown individual from the cultivar 'Sustain', from which the cDNA libraries were made) so it is not possible to know if these sequence variants occur in every genotype or only in some. The primary purpose of cloning the CHR gene from *T. repens* was to obtain a functional gene for expression in transgenic plants, not phylogenetic analysis, and therefore it was not attempted to sequence CHR from a larger number of genotypes. It is also not possible to conclude how many CHR sequence variants exist in any one white clover genotype except that it will be five or more, since five distinct sequence variants were identified from the root cDNA library alone (Figure 2.10). The sequences that were cloned did show two clusters of closely related sequences (Figure 2.11). To determine all CHR sequence variants present in white clover, cloning of large numbers of CHR genomic DNA sequences amplified from a diverse set of primer pairs, or enriched through annealing of genomic DNA to CHR probes, would need to be carried out until no new sequence variants were discovered. Alternatively whole genome sequencing of white clover could be carried out and then searched for CHR sequences. Sequencing of the complete white clover genome from several plant genotypes is planned to occur in 2011 by the Pastoral Genomics Consortium so this question can be answered soon.

CHR amplified successfully from root, leaf and stolon cDNA showing that the gene is expressed, at least minimally, in all of these tissues. Based on cDNA sequencing, at least five different sequence variants were expressed at one time in the root cDNA library. It would be interesting to explore this further to learn what the roles of the different sequence variants are, whether they have different functions and if they are differentially regulated. Functional testing *in vitro* with a range of possible substrates could identify variations in enzyme activity or specificity, complexity that may be exploited by the plant to broaden its metabolic capabilities. Alternatively it is possible that some or all of these sequence variants are simply duplicated genes that have accumulated sequence variations over time and have overlapping functionality and regulation. Given how closely related these genes are in amino acid sequence, compared with the functionally confirmed CHR P10 from soybean (Figure

2.11), it is likely that they are all functionally equivalent and it may be the regulatory elements that vary and allow for differential expression of CHR.

2.4.2 IFS

Full length white clover genomic IFS genes were cloned and found to have two exons with a coding sequence of 1575 bp and 574 aa (Figure 2.2). Aligning all genomic and cDNA clones separated the sequences into 13 sequence variants which varied by SNPs and intron length for the genomic sequences (Figure 2.13).

IFSg_2-12_M-M2 (IFS2_12) has high sequence and amino acid similarity to other publically available IFS genes although they are not as closely clustered as the genes identified in a previous study (Jung et al., 2000). The white clover sequences form two clusters which are distinct from the sequences in *T. repens* and other species from that publication (Figure 2.14). This may be a result of the cloning strategies used to obtain these sequences where primers were designed to one known sequence, soybean IFS1 in the case of Jung et al. (2000), which represents one sequence variant of the gene and amplifies the related sequence variant from other species. Other sequence variants present in the plant genome would not be amplified and cloned, and remain undiscovered. The sequence used to search for IFS genes in the white clover Gene-Thresher database in this project was the white clover IFS1 gene from Jung et al. (2000) and yet the genes cloned from the primers designed to IFS Contig 3 were distinct from white clover IFS1. If another sequence variant had been assembled from the database and used for designing PCR primers, a different set of genes might have been cloned.

The sequence variants that include IFSg_1_08 and IFSg_2_07, which share an intron, are not closely related to any of the IFS cDNA clones (Figure 2.13) which suggests that those genes have no, or low, expression relative to IFSg_2-12, although PCR bias may exist and no dedicated expression analysis has been done that differentiates between sequence variants of IFS.

The sequence variant containing IFSg_1_08 has a deletion of one amino acid in the first exon compared with the others (data not shown). Sequence variants 2-12 and 2-14 vary in

only a few nucleotides and only three amino acids and have the same intron. Sequence variant 2-12 was chosen as the most likely to be an active and expressed gene in the plants.

Three active amino acid residues Ser 310, Lys 375 and Leu 371 (Sawada et al., 2002, Sawada and Ayabe, 2005) are present in the correct position and order and the regions around those residues are conserved among the cloned genes (Figure 2.15) and the diverse public IFS genes (data not shown). This suggests that the sequence variants identified here are functional.

2.4.3 IOMT

Deavours et al. (2006) analysed eight *M. truncatula* IOMT genes for their substrate specificity and investigated which part of the isoflavonoid molecule was being methylated. Deavours et al. (2006) modelled the enzymes in 3D and determined which amino acid residues affected the shape and size of the substrate binding pocket and the substrate specificity. That work found that the most frequent substitutions of active site residues in the MtI7OMT clade of plant OMTs were at those amino acid positions forming the bottom and back wall of the substrate binding pocket. These include a Tyr 25, Tyr 127, Asn 310 and Met 311 (numbered as in MsI7OMT), which were shown previously to be critical residues for either hydrogen bond formation or van der Waals' interactions with phenolic substrates in the structure of MsI7OMT complexed with isoflavanone (Zubieta et al., 2001). Each of those four key amino acids was identical to MsI7OMT in the three sequence variants IOMTg10, 12 and 22 (Figure 2.17). Furthermore, three amino acid residues that are involved in SAM-binding, Lys 253, Trp 259 and catalysis, His 257 (Zubieta et al., 2001) were also all present in all of the sequence variants. This is a good indication that these sequence variants will have similar enzyme activity to MsI7OMT.

All of the cloned white clover IOMT genes were clustered very closely together compared with the eight IOMT genes from Deavours et al. (2006) and were similar to the closely related MtIOMT1, MtIOMT2 and MtIOMT3 genes (Figure 2.16) suggesting that all of the cloned sequence variants have the same enzymatic activity as these *M. truncatula* enzymes. These enzymes had a preference for isoflavonoid rather than flavonoid substrates and methylated 6,7,4'-trihydroxyisoflavone, 7,3',4'-trihydroxyisoflavone, daidzein, genistein,

glycitein and dihydrodaidzein and it is likely that the white clover genes are similarly active on isoflavones and isoflavanones. This also suggests that the white clover genes that are equivalent to MtIOMT4 to 7, which methylate a variety of isoflavanones, isoflavans, coumestans and flavanones, remain undiscovered.

Interestingly MtIOMTs 1 to 3 exist in a gene cluster within a 20kb region of DNA in BAC clone AC146549 (Deavours et al., 2006) which indicates that these three sequence variants arose from gene duplications followed by independent mutagenesis and evolution over time to produce the current differentiated, but similar, sequence variants. It is possible that this evolutionary mechanism gave rise to the large diversity of flavonoid biosynthesis enzymes that is found today.

2.4.4 ANT1

The ANT1 gene cloned from tomato var. ‘Grosse Lisse’ had a nucleotide sequence that was identical to that of the published gene from tomato var. ‘Micro-Tom’ (Mathews et al., 2003) and thus the gene product will have the same functionality. No further sequence analysis was necessary.

2.4.5 Summary

The white clover genes clustered separate from the public gene sequences for CHR (Figure 2.11), IFS (Figure 2.14) and IOMT (Figure 2.16). This shows that the initial sequence that is used to design primers for amplifying more sequence variants in the same or different plant species can affect which, and how many, sequence variants are amplified in that PCR. The primer sequence is specific enough, and the sequence variants variable enough, to amplify only a subset of the biosynthesis enzyme sequence variants, and some sequence variants are likely to remain unamplified and not cloned until different primers are used or a different cloning approach is taken. Thus, while the primary aim of this section of work was to obtain one functional full length gene for overexpression in transgenic plants, some phylogenetic analysis of the sequence variants present in white clover was possible and it indicates that further sequence variants and classes of the three isoflavonoid biosynthesis genes are likely to exist undiscovered in the white clover genome.

The various lines of evidence of sequence similarity with genes with confirmed functionality and the presence of key amino acid residues suggest that the three isoflavonoid biosynthesis genes cloned from white clover are likely to be functional genes. The results from Deavours et al. (2006) showed that, given our current knowledge, sequence similarity cannot yet be used to assign a definitive catalytic function to a gene using bioinformatics alone. Such functionality can only be confirmed by expressing the gene and demonstrating enzyme activity of the protein *in vitro* or *in vivo* in an organism or system that cannot naturally produce the metabolic products of the relevant enzymatic reactions. For this reason functional testing of IFS 2_12 in transgenic tobacco plants has been carried out and is described in the next chapter.

3 Expression of IFS in transgenic tobacco plants

3.1 Introduction

One of the central hypotheses of this thesis is that cloned IFS genes from white clover are both necessary and sufficient for the production of genistein in transgenic tobacco plants expressing these genes. The aim of this chapter is to confirm the enzymatic functionality of the cloned IFS genes by expressing them in a model plant system and detecting the enzyme product, genistein, and its conjugates genistin and genistin malonate, in leaf tissues.

To test this hypothesis, IFS2_12, a putative isoflavone synthase gene from white clover whose cloning was described in the previous chapter, and the functionally confirmed IFS/CHI fusion gene, derived from soybean and alfalfa, respectively (Tian and Dixon, 2006), were transformed into the model plant, tobacco. The genes were transformed by stable transformations and by transient expression in the leaf and the IFS genes were co-expressed with the anthocyanin-inducing transcription factor, ANT1, through the combination of these techniques and through crossing of stable transformants of each gene.

Our knowledge of the mechanisms of isoflavonoid biosynthesis enzyme function is limited, so cloned isoflavonoid biosynthesis genes cannot be confirmed as being functional by sequence analysis alone, as seen by the results of Deavours et al. (2006). The exception is when a cloned gene is identical in sequence to a functionally confirmed gene, as in the case of ANT1 (Mathews et al., 2003, Section 2.4.4). Since white clover transformation is time consuming and inefficient, which severely limits the numbers of transformations achievable in a set period of time, it is beneficial to test the functionality of the cloned genes (Chapter 2) using model systems prior to attempting overexpression of isoflavonoid biosynthesis genes in white clover.

The model plant species tobacco can be transformed at high efficiency and regenerates and grows relatively quickly. It can also be used for transient expression in which an *A. tumefaciens* suspension is infiltrated into the interstitial space in the leaves (Kapila et al., 1997; Wroblewski et al., 2005). The bacteria then independently transform thousands of

individual leaf cells and the infiltrated area of the leaf expresses the transgene for the life of the leaf (Voinnet et al., 2003). The time between transformation and harvesting of leaf material is only seven days and the system potentially allows expression of a large number of transgenes and the simultaneous expression of two or more transgenes. This complements stable transformations which are active for the life of the plant but take much longer to produce and grow to a sufficient size for sampling.

Wild type tobacco does not produce isoflavonoids (Yu et al., 2000) so any biosynthesis of isoflavonoids in transgenic tobacco plants can be attributed to the transgenes. Thus the presence of genistein in transgenic tobacco expressing IFS would confirm the functionality of this gene. Even with an expressed and functional transgene, however, the plants may still not be able to produce or accumulate isoflavonoids, due to a lack of enzymes such as glycosyl transferases to glycosylate the isoflavonoids for transport and storage, or due to a lack of precursor molecules. Tobacco was chosen because Tian and Dixon (2006) transformed tobacco plants with soybean IFS and soybean-alfalfa IFS/CHI and detected isoflavonoids in the transformed plants. Thus the tobacco metabolic background is suitable for production of isoflavonoids.

Yu et al. (2000) found that tobacco plants expressing the soybean IFS enzyme produced about 130 times greater concentration of genistein in flower petals, which were coloured pink by the presence of anthocyanins, compared with leaves which contained no anthocyanins. It was proposed that, in the flower petals, the production of anthocyanins resulted in a higher flux through the first step of the general flavonoid pathway, thereby increasing the concentration of naringenin, a common precursor for anthocyanins, flavonols and isoflavonoids. This may be necessary for producing concentrations of genistein that are large enough to detect, particularly when using HPLC. Thus for functional testing of isoflavonoid biosynthesis genes, in both stable and transient expression, it may be useful or even necessary to activate the anthocyanin pathway in the leaves.

Plants have previously been transformed with transcription factors to produce anthocyanins but often anthocyanins are only produced in normally uncoloured tissues when *two* interacting transcription factors, R2R3 MYB and basic helix-loop-helix (bHLH) proteins are present (Quattrocchio et al., 1998), along with an endogenous WD40 protein (de Vetten et

al., 1997; Larkin et al., 1999; Walker et al., 1999; Hernandez et al. 2004; Grotewold, 2006). The first such example was *A. thaliana* and tobacco transformed with the MYB C1 and/ or bHLH R from maize (Lloyd, 1992). Lloyd (1992) found that C1 transformed alone had no phenotypic effects, R alone intensified anthocyanins in areas where anthocyanins were normally produced, and both C1 and R allowed the production of anthocyanins in areas where anthocyanins were normally never produced, such as the roots, petals and stamens of *A. thaliana*. A single transgene from one of these classes can also interact with an endogenous factor, believed to be the other of the two interacting factors, such as in petunias transformed with the bHLH transcription factor, Lc, whose vegetative tissues turn purple only when high light induces an endogenous factor believed to be a MYB transcription factor (Albert et al., 2009).

Some single transgenes are able to induce anthocyanins in transformed plants, thus making transformation easier than if two genes need to be transformed (Braun and Grotewold, 1999). An R2R3 MYB, Production of Anthocyanin Pigment 1 (PAP-1), was discovered through activation tagging in *A. thaliana* and found to induce anthocyanins in the whole plant in both *A. thaliana* and tobacco (Borewitz, 2000). The levels of pigmentation in *A. thaliana* increased in high light and in response to stress which suggests that PAP1-D was interacting with one or more endogenous bHLH factors that are expressed in all tissues and are upregulated in response to such stresses. Mathews et al. (2003) discovered an R2R3 MYB transcription factor, ANT1, in tomato. This transcription factor on its own was able to induce anthocyanin biosynthesis, modification and transport when expressed in transgenic tobacco plants. If ANT1 requires bHLH and WD40 proteins to activate anthocyanin biosynthesis genes, then those factors are already present and expressed in tobacco plants and the added expression of ANT1 is sufficient to produce anthocyanins.

In this study, the model system that was chosen to perform functional testing of potentially several, and different combinations of, isoflavonoid biosynthesis genes and their sequence variants, was transient expression in tobacco leaves. To increase the likelihood of producing a detectable level of isoflavonoids, tobacco plants were stably transformed to express ANT1 and IFS. These plants were then infiltrated with suspensions of *A. tumefaciens* carrying a vector to express IFS or ANT1 and the isoflavonoid content of the leaves was quantified seven days post-infiltration.

3.2 Materials and Methods

3.2.1 Plant materials

N. tabacum variety ‘Wisconsin 38’ (‘W38’) seed was obtained from stocks held at Agresearch Grasslands (Palmerston North, New Zealand). Seeds were surface-sterilised in bleach (5% available chlorine) for 20 min followed by five washes with sterile water. The seed was germinated and grown on MS0 nutrient agar (Appendix 1) under sterile conditions in a growth room under 16 hour per day of diffuse fluorescent lighting, at 20°C.

3.2.2 Transformation vectors

The Gateway-adapted binary vector pRSh1 was used to transform genes into tobacco. This contains, within the left and right borders of the transferred DNA, the BAR gene, that provides resistance to ammonium glufosinate herbicide (Thompson et al., 1987), and a Gateway cloning site where the gene of interest is inserted. Both genes are driven by a Cauliflower Mosaic Virus (CaMV) 35S promoter. The genes white clover IFS2_12 (Section 2.4.2), soybean-alfalfa IFS/CHI (Section 2.3.7), and tomato ANT1 (Section 2.4.4) were cloned into pRSh1 (Appendix 2) and used for stable and transient transformation of tobacco. See Appendix 2 for vector maps.

The vector pRS13, based on pRSh1, contains the Green Fluorescent Protein (GFP) and β -glucuronidase (GUS) genes driven by CaMV 35S promoters and was used as a control to assess whether infiltration was successful (Appendix 2). The binary vector pHZbar was used as a vector control, containing only the BAR gene within the left and right borders. The empty vector pRSh1 contains the ccdB gene which is lethal to *E. coli*. As this gene may adversely affect the plant cells, it was not used as a vector control (Appendix 2). The binary vector pBin61 (Bendahmane et al., 2000) contained the P19 protein, from tomato bushy stunt virus, which is a suppressor of post transcriptional gene silencing (Voinnet et al., 2003). When Voinnet et al. (2003) co-infiltrated this into *Nicotiana benthamiana* leaves with other vectors expressing GFP, the total expression of the GFP protein was increased compared with when only the GFP-expressing vector was used.

3.2.3 Transformation and culture of *Agrobacterium tumefaciens*.

Competent *A. tumefaciens* 'GV3101' cells (provided by D. Maher) were slowly thawed in an ice bath and about 1 µg of plasmid DNA added. The cells were snap frozen in liquid nitrogen for a few seconds and then incubated in a 37°C water bath for 5 min. Following this, 0.9 ml of LB medium (Appendix 1) was added to the tube and incubated at 28°C for 2 to 4 h with gentle shaking. Cells were then centrifuged, the supernatant poured off and the cells resuspended in 100 µl LB medium. Cells were spread on a YM plate (Appendix 1) containing 100 µg ml⁻¹ spectinomycin and incubated at 28°C. Transformed colonies appeared after 2 to 3 d.

Individual colonies were picked and grown overnight in 5 ml MGL broth (Appendix 1) at 28°C and 200 rpm shaking. To preserve the transformed bacteria, 800 µl of bacterial suspension mixed with 200 µl 100% glycerol were stored at -80°C. The remaining medium was centrifuged for 5 min at 3220 g to pellet the bacteria and the plasmid DNA was extracted for verification using a QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions except for performing an extended time (15 min) lysis step.

3.2.4 Stable transformation of *Nicotiana tabacum*

The protocol to transform and regenerate tobacco is based on that of Horsch et al. (1985). Stable transformations were performed separately with the genes ANT1, TrIFS2_12 and IFS/CHI, all cloned in the vector pRSh1.

The *A. tumefaciens* strain carrying the desired pRSh1-based vector was grown overnight in 20 ml liquid MGL medium (Appendix 1) containing 100 µg ml⁻¹ spectinomycin at 28°C with shaking at 200 rpm. The bacteria were pelleted by centrifugation and resuspended in 5 ml of sterile 10 mM MgSO₄.

Under sterile conditions in a laminar flow hood, leaves of approximately 2 to 5 cm width were cut from a single sterile tobacco plant and sliced into 0.7 cm by 2 cm strips using a sharp scalpel. The leaf strips were immersed in the suspension of *A. tumefaciens* for 5 min

and then blotted on sterile filter paper. The dissected leaves were placed onto non-selective, shoot-inducing MS101 agar containing no antibiotics or herbicides (Appendix 1).

Plates were sealed with plastic film and explants were cultured for 2 d in a growth room at 25°C under artificial lighting for 16 h per day. They were then transferred onto a fresh selective MS101 agar plate containing 300 $\mu\text{g ml}^{-1}$ timentin antibiotic (TIM, ticarcillin) and 2.5 $\mu\text{g ml}^{-1}$ ammonium glufosinate (DL-phosphinothricin) herbicide and returned to the growth room. From then onwards, explants were transferred onto fresh selective MS101 agar plates every 2 weeks. Callus and shoots grew within 2 to 6 weeks and shoots were transferred onto selective MS0 agar media. MS0 is a root-inducing medium that contains no plant hormones (Appendix 1). Callus and shoots of tissue successfully transformed with, and expressing, ANT1 were coloured dark red due to the accumulation of anthocyanins and were thus very easy to identify visually and isolate from light green callus or shoots that were not transformed or were transformed but not expressing ANT1. For other vectors, no colour selection was possible and vigorously growing shoots were selected for further analysis.

Shoots that produced roots were transferred into sterile peat plugs in large tissue culture containers. After about two weeks, roots emerged from the sides of the peat plug and plants were moved to a containment greenhouse and, after a week of acclimatisation, were planted into potting mix in PB3 or PB5 planter bags. The plants subsequently remained in the containment greenhouse and were destroyed by autoclaving when no longer required. The tobacco plants were labelled by species, transgene and a number (eg Nt ANT1 01) and were assigned a unique number (eg GH 23726) once transferred to the greenhouse. In following text the tobacco regenerants transformed with white clover IFS2_12, soybean-alfalfa IFS/CHI and tomato ANT1 are referred to as Nt-IFS2_12, Nt-IFS/CHI and Nt-ANT1 plants respectively.

3.2.5 Plant DNA extraction

DNA was extracted from plants using two methods. The first used DNAzol (Invitrogen) where approximately 100 mg FW of leaf tissue was placed in a 2 ml screw-top microtube with 400 μl of DNAzol, 1 μl RNase A (Invitrogen) and a 3mm stainless steel ball. The sample was shaken in a FastPrep FP120 mixer mill twice for 30 s at a speed of 6.5 m s^{-1} and

incubated at room temperature for 5 min. The sample was centrifuged at 16600 g for 10 min and 300 µl of supernatant transferred to a new microtube. Ethanol (225 µl) was added and mixed and the mixture transferred to a glass filter spin column (DNeasy Plant Mini Kit, Qiagen) and centrifuged at 6000 g for 4 min. The column was washed with 500 µl Buffer AW (DNeasy Plant Mini Kit, Qiagen) and centrifuged at 6000 g for 1 min. The column was then washed twice with 500 µl of 95% v/v ethanol and centrifuged at 6000 g for 1 min. The column was washed again with 500 µl of 95% v/v ethanol and centrifuged at 16600 g for 2 min and air-dried for 5 min. The DNA was eluted from the column into a clean microtube by adding 100 µl of 65°C 10 mM Tris-HCL pH8.0 to the glass filter, incubating for 5 min at room temperature and centrifuging at 6000 g for 1 min. The elution was repeated for a total eluate of 200 µl.

The second plant DNA extraction method used a REDExtract-N-Amp Plant PCR Kit (Sigma-Aldrich). A small (5 to 7 mm diameter) leaf explant was incubated with 100 µl of Extraction Solution at 95°C in a microtube. After 10 min, 100 µl of Dilution Solution was added and mixed. This solution was used as PCR template.

3.2.6 PCR confirmation of tobacco transformation

Genomic DNA was extracted (Section 3.2.5) from the tobacco regenerants and tested for the presence of the inserted transgene using the standard PCR protocol (Section 2.2.5). PCR products were resolved by agarose gel electrophoresis (Section 2.2.8) and a band of the predicted size indicated the presence of the transgene in the plant genomic DNA. Genomic DNA was also extracted from five Nt-IFS/CHI plants using a REDExtract-N-Amp Plant PCR Kit (Section 3.2.5) and PCR tested using the XNAP protocol (Section 3.2.7). The primers used and the expected product sizes are shown in Table 3.1. Primer sequences are presented in Appendix 4.

Table 3.1 Primers used to test for transgenes in tobacco (see Appendix 4 for primer sequences).

Gene	Forward primer	Reverse primer	Expected product size (bp)
IFS	pRSh1 F2	IFS RT_R2	1463
ANT1	pRSh1 F2	ANT1 R+attB2	1101
IFS/CHI	pRSh1 F2	IFS 5' INT R	144

DNA from four Nt-IFS2_12 regenerants was tested, one of which had two shoots tested, along with two wild type tobacco plant negative control DNA samples, one sample of pRSh1+IFS2_12 plasmid DNA as a positive control and a no-DNA blank.

DNA from ten Nt-ANT1 regenerants was tested, one of which (GH24366) exhibited areas of green and red coloured leaf tissue. Both of these phenotypes were PCR tested for the transgene. As controls, two wild type tobacco plant negative control DNA samples, one pRSh1+IFS2_12 plasmid DNA positive control and a no-DNA blank were also amplified.

DNA from five Nt-IFS/CHI regenerants, one 'W38' wild type tobacco negative control plant, one pRSh1+IFS2_12 plasmid DNA positive control and a no-DNA blank were PCR tested using the XNAP PCR protocol (Section 3.2.7).

3.2.7 XNAP PCR

Plant genomic DNA extracted with DNAzol was PCR tested using the Standard PCR protocol (Section 2.2.5). Genomic DNA extracted using the REExtract-N-Amp Plant PCR Kit was PCR tested as described below using the PCR ReadyMix included in the kit. The reaction contained 10 µl of REExtract-N-Amp PCR ReadyMix, 4 µl of DNA extract and 0.4 µM of each of the forward and reverse primers, in a total volume of 20 µl. PCR reactions were performed in a thermocycler (MJResearch PTC200, ABI9700, ABI2720 or BioRad iCycler) running the following temperature profile: 94°C for 3 min, 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 120 s, followed by 72°C for 10 min.

3.2.8 Transient expression of transgenes by leaf infiltration of *Nicotiana tabacum*.

The leaf infiltration method is an adaptation of the method document “Transient expression of *Agrobacterium* in *Nicotiana benthamiana*” from the Sainsbury Laboratory (John Innes Centre, Norwich, UK, personal communication) which was based on a method by Kapila (1997).

An *A. tumefaciens* ‘GV3101’ strain carrying the desired vector was grown overnight in 20 ml liquid MGL medium (Appendix 1) containing the appropriate antibiotics (100 µg ml⁻¹ spectinomycin or 50 µg ml⁻¹ kanamycin) at 28°C with shaking at 200 rpm. The bacteria in the broth were pelleted by centrifugation and the cells resuspended in 10 ml of sterile infiltration buffer (10 mM MgCl₂, 10 µM acetosyringone) which was diluted ten-fold in the final experiment. The bacterial suspension was incubated for 2 to 3 h at room temperature. If more than one gene was to be co-expressed, separate bacterial suspensions were mixed together before infiltration.

Infiltration of tobacco leaves was performed by filling a 1 ml sterile plastic syringe with the appropriate infiltrate and making a very small cut in the underside of a leaf with the tip of a wide 16 gauge needle. The tip of the syringe (without a needle fitted) was pressed onto the underside of the leaf over the hole while a fingertip was pressed onto the opposite side of the leaf over the syringe opening and the cut in the leaf. The infiltrate was then gently and steadily injected into the interstitial spaces of the leaf until there was an increase in the pressure required to infiltrate further and the visible infiltration of liquid slowed down. The optimal tobacco tissue to infiltrate was the first few large leaves of approximately 15 cm length which had a rounded shape and developed approximately three weeks after transplanting small plants from peat plugs into pots in the greenhouse. The resistance to infiltration was greater in small young leaves and also in older leaves which had a narrower shape with more pointed leaf tips, resulting in smaller infiltration areas and more infiltration sites being required to cover a defined area. The optimal area to infiltrate within a leaf was the flat tissue between veins since vascular tissue resisted the flow-through of infiltrate.

Infiltrated areas were left for between one and four weeks (seven days unless mentioned) to give time for the transgene to be expressed and for metabolites to accumulate. Infiltrated areas were then dissected from the leaves with a scalpel, freeze-dried and weighed prior to isoflavonoid extraction.

3.2.9 Transient expression experiments

1. The first infiltration experiment was carried out on young Nt-ANT1 and Nt-IFS2_12 regenerants when they were still less than 10 cm tall. All plants had four infiltrations into one leaf, with each infiltration in separate areas of the leaf. All plants had no-infiltration, pHZbar empty vector and pRS13 control. The pRS13 vector contains green fluorescent protein (GFP) and β -glucuronidase (GUS) genes. ANT1 stable transformants were also infiltrated with IFS while IFS stable transformants were also infiltrated with ANT1. All infiltration solutions were a 1:1 mixture of the desired vector and the P19 Trans-1-RS-86 vector.
2. The second experiment was a repeat of the first at a smaller scale six weeks later when plants and leaves were larger and more robust. This was carried out to test infiltration on larger leaves and also to test if infiltration of water or infiltration buffer without *A. tumefaciens* was harmful to the leaf. Five spatially separated infiltrations were performed on each leaf. The solutions infiltrated were: P19+pHZbar, P19+IFS or P19+ANT1, P19+pRS13, infiltration buffer without *A. tumefaciens*, and water. Two Nt-ANT1 plants and two Nt-WT plants were infiltrated.
3. The third experiment tested the effect of *A. tumefaciens* concentration in the infiltration solution on the infiltrated areas of wild type tobacco. After the *A. tumefaciens* were grown overnight, they were pelleted and resuspended in infiltration buffer and mixed in the following combinations: P19+ANT1, P19+pRS13 as well as ANT1 and pRS13 alone. A 2x serial dilution of the *A. tumefaciens* suspension combinations was performed using infiltration buffer as the diluent. This produced six dilutions ranging from 1/1 to 1/32th of the original *A. tumefaciens* suspension combination. Each leaf that was used had one combination infiltrated with six spatially separate infiltration areas for the six dilutions.
4. The fourth experiment was a follow up of the third experiment to determine the lower limit of effective *A. tumefaciens* dilution. This experiment used only ANT1

infiltrates in a 4x serial dilution with infiltration buffer so that the dilutions were: 1/1, 1/4, 1/16, 1/64, 1/256 and 1/1024 in wild type tobacco.

5. The fifth experiment was carried out to test whether infiltration with a precursor allowed isoflavonoid production at levels detectable by HPLC. Bronze-coloured Nt-ANT1 leaves were infiltrated in six places with either infiltration solution without any *A. tumefaciens* (plant 5P01969) or with *A. tumefaciens* carrying IFS/CHI (plant 5P01967) at a 1/10 dilution. After seven days, the infiltration sites were visually assessed and one infiltration site for each condition was harvested along with an uninfiltrated area of leaf. The remaining five infiltration sites were additionally infiltrated with 2.5 mg ml⁻¹ naringin (naringenin rhamnoglucoside) in a 5% v/v ethanol solution and harvested after different incubation times – 0 min, 15 min, 30 min, 60 min and 120 min. The samples were extracted and analysed by HPLC.
6. The sixth experiment used the optimised infiltration and analysis protocols to test Nt-ANT1 and NT-IFS/CHI plants infiltrated with 1/10 dilutions of infiltration solutions containing the genes that the plants are not stably transformed with: Nt-ANT1 plants were infiltrated with pHZbar, IFS2_12 or IFS/CHI while Nt-IFS/CHI plants were infiltrated with pHZbar or ANT1. One half of a leaf was infiltrated for each sample. Seven days post-infiltration, uninfiltrated and infiltrated leaf tissue samples were freeze-dried whole, ground in a Fast Prep mixer mill and isoflavonoids were extracted three times with acetone (Section 3.2.12). A no-leaf-sample extraction was also carried out to test for any possible genistein contamination that could be introduced by the extraction method. LCMS analysis was carried out (Section 3.2.14) with 10 µl sample injection and all samples were interspersed with blank injections of water to minimise carryover contamination. The 61 min run protocol was used and the LCMS was set to scan for MS1 ions of m/z 271, 433 and 519, MS2 fragments of m/z 271 and MS3 fragments of m/z 153 (Section 3.2.14).

3.2.10 Crossing of stable tobacco transformants

Once mature, the Nt-IFS2_12, Nt-ANT1 and Nt-IFS/CHI plants from section 3.2.4 that produced flowers at the same time were crossed as well as allowing mature flowers to self-fertilise. Crossed flowers were labelled with tags identifying the pollen donor and mature seed pods were collected and stored at 4°C. Seedlings were grown and leaf tissue collected

from three to six young plants from each cross. The original Nt-ANT1 5P01963 plant and three 'W38' wild type control plants were also sampled. Leaf tissue was freeze-dried, ground and isoflavonoids extracted as described in section 3.2.12 and analysed by LCMS (Section 3.2.14).

3.2.11 Testing of GFP and GUS expression

Leaves infiltrated with pRS13 were tested for the expression of GFP and GUS. Leaves were examined for the presence of GFP under a fluorescent microscope with GFP filters and assessed visually.

The same leaves were stained for the presence of GUS by incubating tissue in a 6-well microplate with a staining solution composed of 0.1 M NaPO₄ pH 7.0, 10 mM EDTA, 0.1% w/v Triton X-100, 1 mM K₃Fe(CN)₆ and 2 mM X-Gluc (5-Bromo-4-chloro-3-indoxyl-beta-D-glucuronide cyclohexylammonium salt) (Jefferson et al., 1986). The samples were incubated at 37°C overnight and then washed two times with 50% v/v ethanol for 12 h per wash. GUS expression, evident by blue colouration, was assessed by visual inspection.

3.2.12 Isoflavonoid extraction

Samples were placed in pre-weighed 2 ml screw top microtubes (Sarstedt) pre-loaded with five 3 mm steel balls or a 200 µl volume of 1 mm ceramic beads, freeze-dried whole (Flexi-Dry™ µP freeze drier), weighed and ground in a Fast Prep FP120 mixer mill twice for 30 s at 4 m s⁻¹.

For some white clover samples (Chapter 4) plant tissue was frozen in liquid nitrogen and ground by hand using a mortar and pestle. Alternatively the samples were placed in stainless steel grinding jars with a 19 mm stainless steel ball, all cooled in liquid nitrogen, and placed in a MM301 mixer mill (Retsch) and shaken at 30 Hz for 30 s. Frozen, ground plant tissue was freeze-dried and approximately 50 mg of dry tissue was weighed out into a 2 ml microtube. The exact weight was noted.

Isoflavonoids were extracted from the sample by the addition of 1 ml of 80% v/v methanol containing 0.1% v/v acetic acid and mixing the sample in an Eppendorf Comfort

Thermomixer at room temperature with shaking at 1400 rpm for 30 min. The sample was centrifuged at 16600 g for 10 min and the supernatant decanted into a new tube. The sample was placed at -20°C for 30 min then centrifuged again. The supernatant (700 µl) was transferred to an HPLC or LCMS autosampler vial, sealed and stored at -20°C prior to analysis.

In the sixth experiment a higher concentration of isoflavonoids was obtained by extracting a greater amount (approximately 150 mg DW) of dry tissue three times with 1 ml acetone for 30 min, centrifuging at 16600 g for 1 min, pooling the acetone extract and evaporating it off under nitrogen gas before the extract was dissolved in 200 µl of 80% v/v methanol containing 0.1% v/v acetic acid.

3.2.13 Quantification of isoflavonoids using high performance liquid chromatography (HPLC)

Krenn et al. (2002) described a method for extracting, separating using HPLC and detecting (using a photodiode array) genistein, daidzein, biochanin A and formononetin from red clover which was the basis of the method used for this study. However, a more sensitive detection method that also allows better identification of compounds is mass spectrometry (MS). Wu et al. (2003) developed a method to identify and quantify isoflavonoids in four *Trifolium* species including red and white clover using high-performance liquid chromatography combined with ultraviolet and electrospray ionisation mass spectrometric detection (HPLC-UV-ESI-MSD). This method was used in the latter part of this work and was essential for measuring low concentrations of isoflavonoids in tobacco.

For HPLC, samples (20 µl) were analysed on a Shimadzu HPLC (SCL-10Avp system controller, LC-10ATvp pump, FCV-10ALvp quaternary valve, DGU-14A degasser, SIL-10ADvp autosampler, CTO-10Avp oven, RF10Axl fluorescence detector and SPD-M10Avp photodiode array (PDA) detector) with a Phenomenex Luna C18(2)-HST column (100 x 2.0mm, 2.5µm particle size; Phenomenex, Torrance, CA) running 1 ml min⁻¹ of water/acetonitrile acidified with 0.1% v/v formic acid. The linear solvent gradient profile was: 0 min 20% v/v acetonitrile, 35 min 37% v/v acetonitrile, 45 min 100% acetonitrile, 50 min 100% v/v acetonitrile, 51 min 20% v/v acetonitrile, 61 min 20% v/v acetonitrile.

HPLC data were analysed using Shimadzu VP7 software, max plots of isoflavonoid peaks were identified and quantified by comparison with biochanin A, genistein and formononetin standards as well as a standards library for their absolute and relative elution time, PDA spectrum and fluorescence.

3.2.14 Liquid chromatography with mass spectrometric analysis (LCMS)

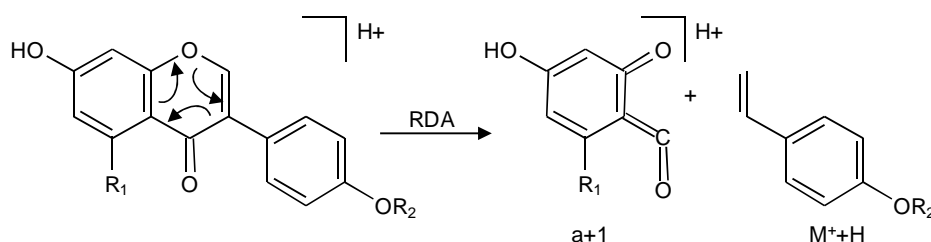
The instrument consisted of two JASCO X-LC 3085PU high pressure pumps (JASCO International Co. Ltd., Tokyo, Japan), and a HTS PAL autosampler with a 25 µl syringe and a 15k psi injection valve (CTC Analytics AG, Zwingen, Switzerland), coupled to a Thermo LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Thermo Finnigan Xcalibur software (version 2.1) was used for data acquisition and processing. A 20 µl injection of sample was made onto a Phenomenex Luna C18(2)-HST column (100 x 2.0mm, 2.5µm particle size; Phenomenex, Torrance, CA). The HPLC solvents used were; solvent A = 0.1% v/v formic acid in MilliQ® water; solvent B = 0.1% v/v formic acid in Acetonitrile (LiChroSolv grade, Merck, New Zealand). The flow rate was 400 µl min⁻¹ and the solvent gradient used was: 0 min 5% v/v acetonitrile. 6 min 5% v/v acetonitrile, 11 min 10% v/v acetonitrile, 26 min 17% v/v acetonitrile, 31 min 23% v/v acetonitrile. 41 min 30% v/v acetonitrile. 45 min 50% v/v acetonitrile. 52 min 97% v/v acetonitrile. 59 min 97% v/v acetonitrile. 61 min 5% v/v acetonitrile. 65 min 5% v/v acetonitrile. The acetonitrile concentration changed linearly between timepoints.

Mass spectra were acquired using ESI in positive mode with a spray voltage of +4.5 kV, an ion source capillary temperature of 275 °C and the nitrogen sheath, auxiliary, and sweep gas flow rates set to 20, 5, and 0 (arbitrary units), respectively. The first 3 minutes of flow from the HPLC were diverted to waste. The mass spectrometer was programmed to scan from 230-1500 m/z (MS1 scan) and then sequentially perform product ion scans for selected masses on the desired different combinations of PC and PD masses (*m/z* 271, 433, 519) with isolation windows for each selected *m/z* value of 2.0 mass units and a fragmentation collision energy value of 35%.

LCMS peaks were identified and quantified by comparison with biochanin A, daidzein, genistein and formononetin standards (Sigma) for their elution time, mass spectrum and

retro-Diels-Alder (RDA) fragment masses of collision energy fragmentations performed in the trap with helium. The mass spectral data contained the mass of the intact compound, the compound with the loss of the malonyl and/or the glycosyl groups, if present, and RDA fragments produced by fragmentation of the aglycone. Quantification was carried out by creating a calibration curve of chemical standards of known concentration, using MS ion count peak areas.

Genistein aglycone, genistin and genistin malonate masses are listed in Figure 1. When retained by the ion trap of the mass spectrometer and fragmented, the glycan residue is removed leaving the aglycone as an MS2 fragment of m/z 271. When this is fragmented further, the aglycone is cleaved to produce two RDA fragments (Figure 1). An MS1 ion or MS2 fragment of m/z 271 and an MS3 fragment of m/z 153 is confirmation that the compound contains genistein. Apigenin also produces a m/z 271 ion but no RDA fragments at m/z 153 (Kuhn et al., 2003).



Name	R ₁	R ₂	a+1	M ⁺ +H	M+malonate ⁺ +H	M+glucoside malonate ⁺ +H
genistein	OH	H	153	271	433	519
daidzein	H	H	137	255	417	503
biochanin A	OH	CH ₃	153	285	447	533
formononetin	H	CH ₃	137	269	431	517

Figure 3.1 The structures of four isoflavone aglycones and their MS-MS fragment pathway (RDA) with molecular ions as their precursors. Based on Kuhn et al. (2003).

Quantification of the peaks was based on ion counts of fragmented m/z 271 ions, so the concentrations and weights of genistein and its glycosides were measured in “genistein aglycone only” units. In other words, the weights do not include the added weights of the glucoside or glucoside-malonate molecules. The true weight of genistin, including the

glucose, in the sample can be calculated by multiplying the measured weight by $433/271 = 1.60$ times. Likewise the genistin malonate results can be multiplied by $519/271 = 1.92$ times to obtain the true weight. MS2 ion count peak area was compared with that of known aglycone quantity standards to quantify the isoflavonoid content in the sample.

3.3 Results

3.3.1 Stable tobacco transformants expressing isoflavonoid biosynthesis genes

Stable transformations of two isoflavonoid biosynthesis genes in tobacco were successful, leading to plant regeneration in tissue culture, the production of roots on regenerated shoots and eventual transfer into soil and the greenhouse. The primary transformants that were tested and used were four Nt-IFS2_12 plants and five Nt-IFS/CHI plants.

The PCR reaction testing Nt-IFS2_12 regenerants (Section 3.2.6) produced the predicted 1463 bp band in the positive control and sample GH 23722 and no band in the other samples or in the negative controls or the blank (Figure 3.2). This confirms GH 23722 as transgenic and the other plants as untransformed.

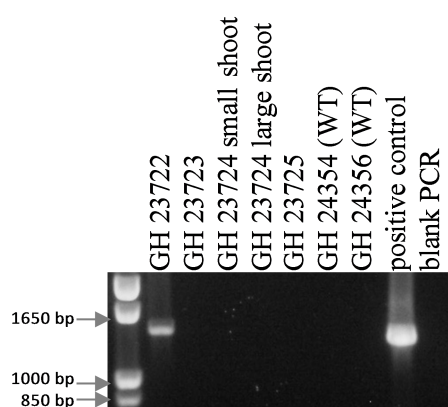


Figure 3.2 Test of Nt-IFS2_12 plants for presence of IFS2_12 transgene.

The PCR reaction testing NT-IFS/CHI regenerants (Section 3.2.6) produced the predicted 144 bp band in the positive control and in samples NT-IFS/CHI 5P02034, 5P02035, 5P02036 and P05 (not transferred to greenhouse yet) and no band in P02, the negative control or the blank (Figure 3.3). This confirms 5P02034, 5P02035, GH 5P02036 and P05 as transgenic and P02 as untransformed.

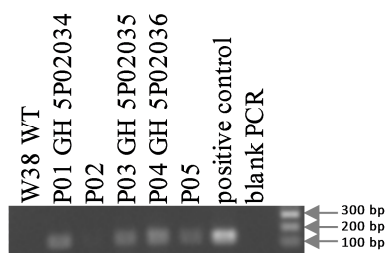


Figure 3.3 Test of Nt-IFS/CHI plants for presence of IFS/CHI transgene.

The Nt-IFS2_12 and Nt-IFS/CHI plants were phenotypically normal, healthy green plants. The Nt-ANT1 plants were phenotypically distinct, as described in the next section.

3.3.2 Stable tobacco transformants carrying ANT1

Ten Nt-ANT1 plants regenerated and were used in subsequent experiments. The PCR reaction testing Nt-ANT1 regenerants (Section 3.2.6) produced the predicted 1101 bp band in the positive control and all test samples, and no band in the negative controls or the blank (Figure 3.4). This confirms all tested Nt-ANT1 plants as transgenic including both red and green leaf areas from plant GH24366.

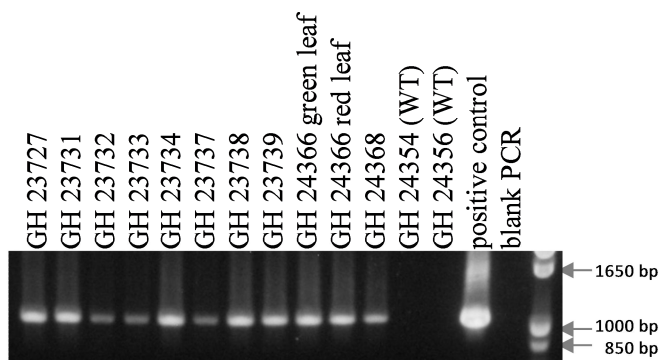


Figure 3.4 Test of Nt-ANT1 plants for presence of ANT1 transgene.

The Nt-ANT1 plants had a range of colour and form phenotypes (Figure 3.5). Some remained green or were phenotypically chimeric with both green and red sectors of leaves. Green regenerants (Figure 3.5 A) were visually detected and mostly discarded so that only bronze to purple coloured plants were used in subsequent leaf infiltration experiments. One phenotype comprised plants with leaves that had a speckled red appearance. These had otherwise normal growth and form but had small areas of light red on the leaves surrounded by areas with less intense red colour or green (Figure 3.5 B). Some plants were coloured a uniform light bronze and these also grew normally (Figure 3.5 C). Many plants had intermediate colour levels and appeared a uniform bronze colour. They had a growth habit, form and vigour identical to wild type plants and appeared to be completely unaffected by the anthocyanin production (Figure 3.5 D). These plants were selected for infiltration and further analysis. Among the coloured plants, some were coloured dark purple and these were physically abnormal. They produced smaller, slightly abnormally shaped leaves, grew more slowly and produced several crooked shoots instead of one straight, apically dominant, shoot (Figure 3.5 E). Another type was coloured in the leaf areas between veins while the areas around the veins remained green (Figure 3.6 C). These plants grew very slowly, were prone to wilting and did not grow very tall.

The bronze coloured plants had red vasculature and dark red root tips with colourless young roots that turned red over time (Figure 3.7). They grew normally and eventually produced flowers that were more intensely pink and purple coloured than the light pink wild type, and seed that were darkly coloured compared with the light brown colour of wild type seed.

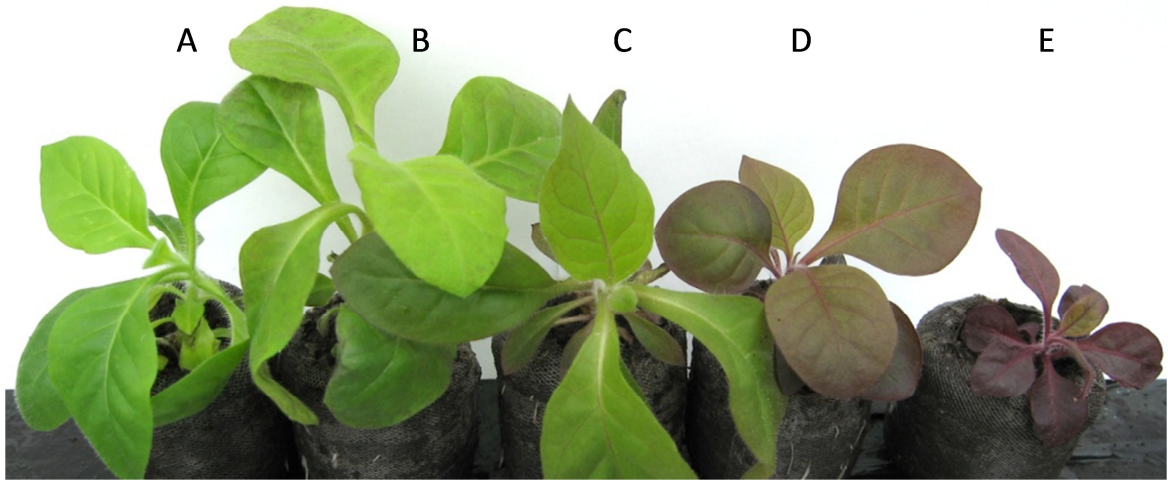


Figure 3.5 Nt-ANT1 regenerants showing the diversity of colour phenotypes. A, green. B, red-speckled green. C, light bronze, D, bronze. E, purple.

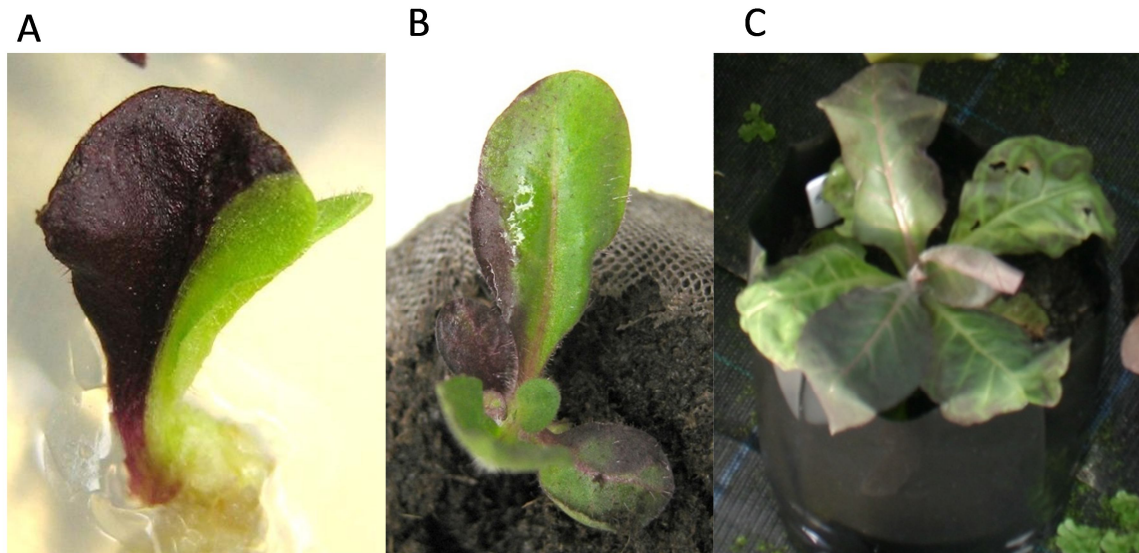


Figure 3.6 Nt-ANT1 regenerants showing phenotypes of both green and coloured (purple) areas in the leaves. A, leaf-like tissue regenerating from callus in tissue culture. B, plant with purple and green leaf areas. C, plant with bronze leaves with green areas around the main leaf vasculature.

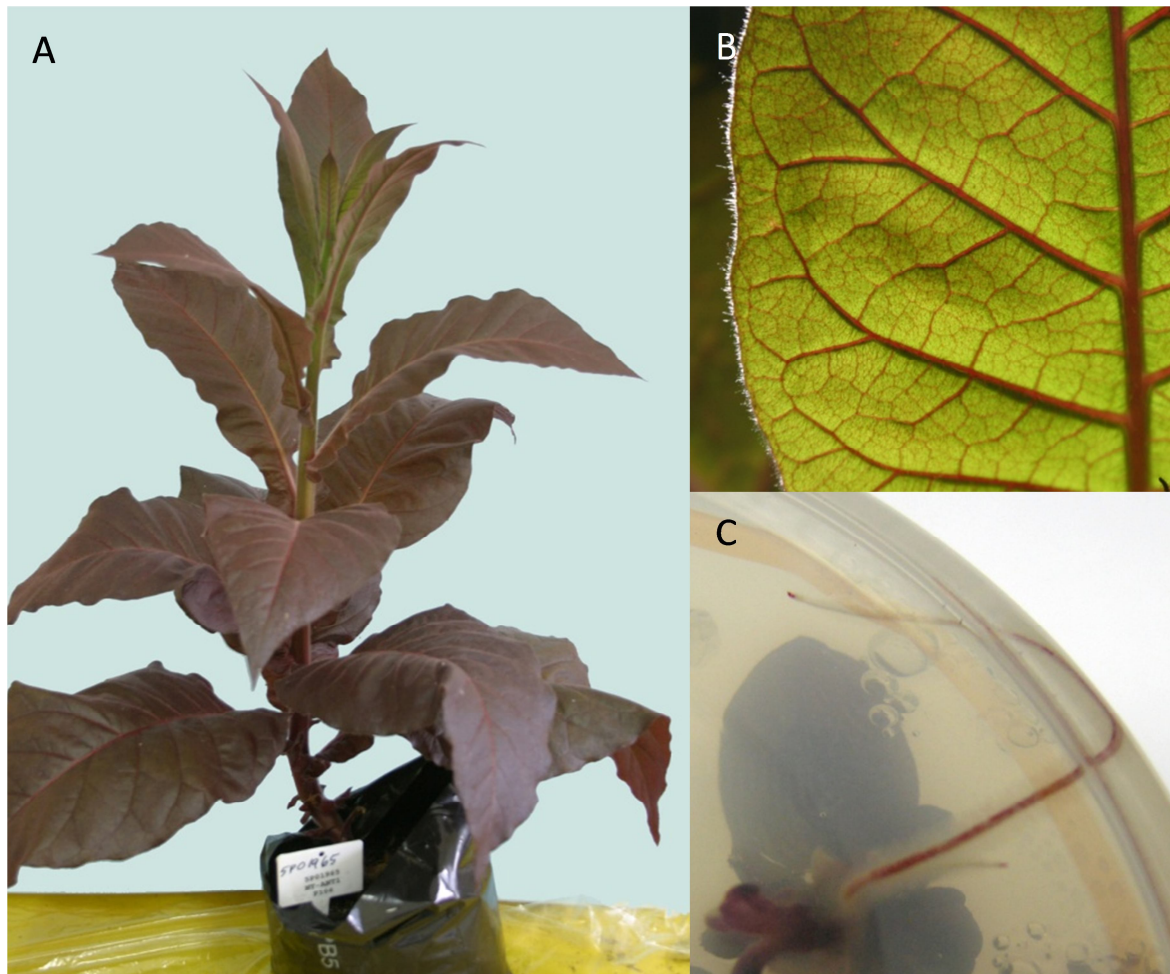


Figure 3.7 Phenotype of bronze tobacco Nt-ANT1 transformants. A, mature plant. B, back lit image of a leaf showing the red vasculature. C, image of plant growing on an MS0 agar plate from underneath showing red root tip (top left), colourless young root and red older root (lower half).

Stable transformation of *N. benthamiana* with ANT1 was also carried out, since this species was reputed to be easier to infiltrate and work with. Unfortunately the coloured transformants were very stunted and misformed and grew very slowly, so work with this species was discontinued.

3.3.3 Transient expression in tobacco through leaf infiltration

In the first infiltration experiment in young Nt-ANT1 and Nt-IFS2_12 regenerants (Section 3.2.9), the infiltrated areas collapsed and went limp within days, so the leaves were harvested five days post-infiltration. A trial plant that was infiltrated only with red food colour (Allura Red, Hansells) in water also became limp (data not shown). Uninfiltrated leaves and uninfiltrated areas within leaves remained turgid and healthy. The effect on the leaves was localised to the infiltrated areas. Figure 3.8 shows representative examples.



Figure 3.8 Leaf infiltration in young tobacco leaves. ANT1 transformants infiltrated with: A, IFS2_12. B, pHZbar. C, pRS13. D, not-infiltrated.

Leaves infiltrated with pRS13 were visually tested for the presence of GFP (Section 3.2.11) which was clearly evident in cells within the infiltrated areas and absent in non-infiltrated areas (Figure 3.9).

Staining for the presence of GUS (Section 3.2.11) resulted in no blue colour being visible in any areas – no GUS activity was detected (data not shown).

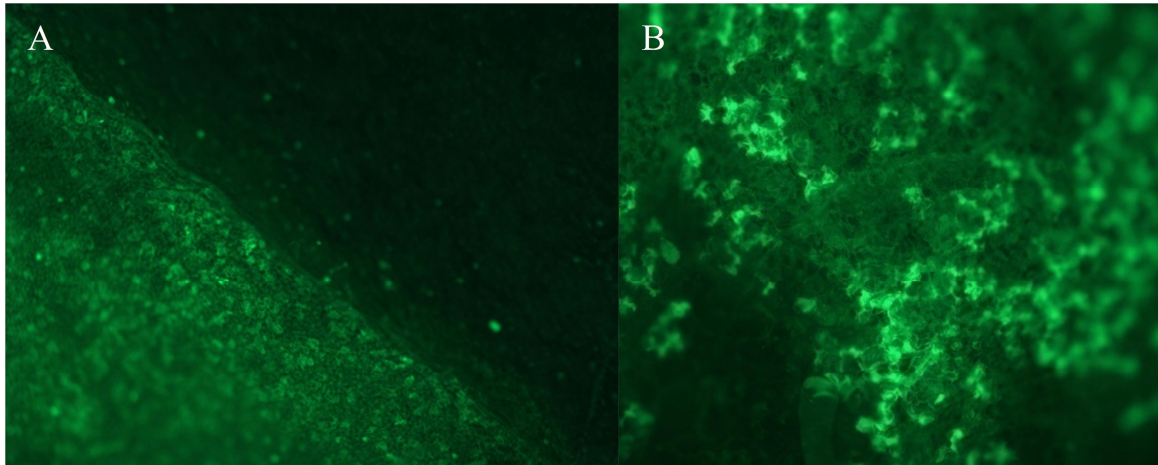


Figure 3.9 GFP expression in leaves infiltrated with pRS13, first infiltration experiment. A, infiltrated (lower left) and non-infiltrated (upper right) areas of a leaf with a diagonal vein as the boundary between them. B, higher magnification view showing that not all cells expressed GFP.

In the second experiment (Section 3.2.9), for both wild type and Nt-ANT transgenic tobacco plants, the infiltrated areas of the larger leaves did not collapse as had the younger leaves in the first experiment. Both the infiltration solution without *A. tumefaciens* and the water infiltrations had no visible negative effects on the leaves of the plant – the infiltrated areas remained turgid and green (data not shown but the infiltrated areas looked the same as Figure 3.11 G). The three areas infiltrated with *A. tumefaciens* were negatively affected and turned yellow and necrotised over time (data not shown but the infiltrated areas looked the same as Figure 3.11 H). Leaves were harvested 16 days later and areas infiltrated with pRS13 were analysed for GFP which was found to be present in the infiltrated areas and absent in the non-infiltrated areas of all plants (Figure 3.10). HPLC analysis was carried out on isoflavonoid extracts of the leaf areas with infiltration solution without *A. tumefaciens*, P19+pHZbar, P19+IFS and P19+ANT1. No isoflavonoids were detected in any of the samples (data not shown).

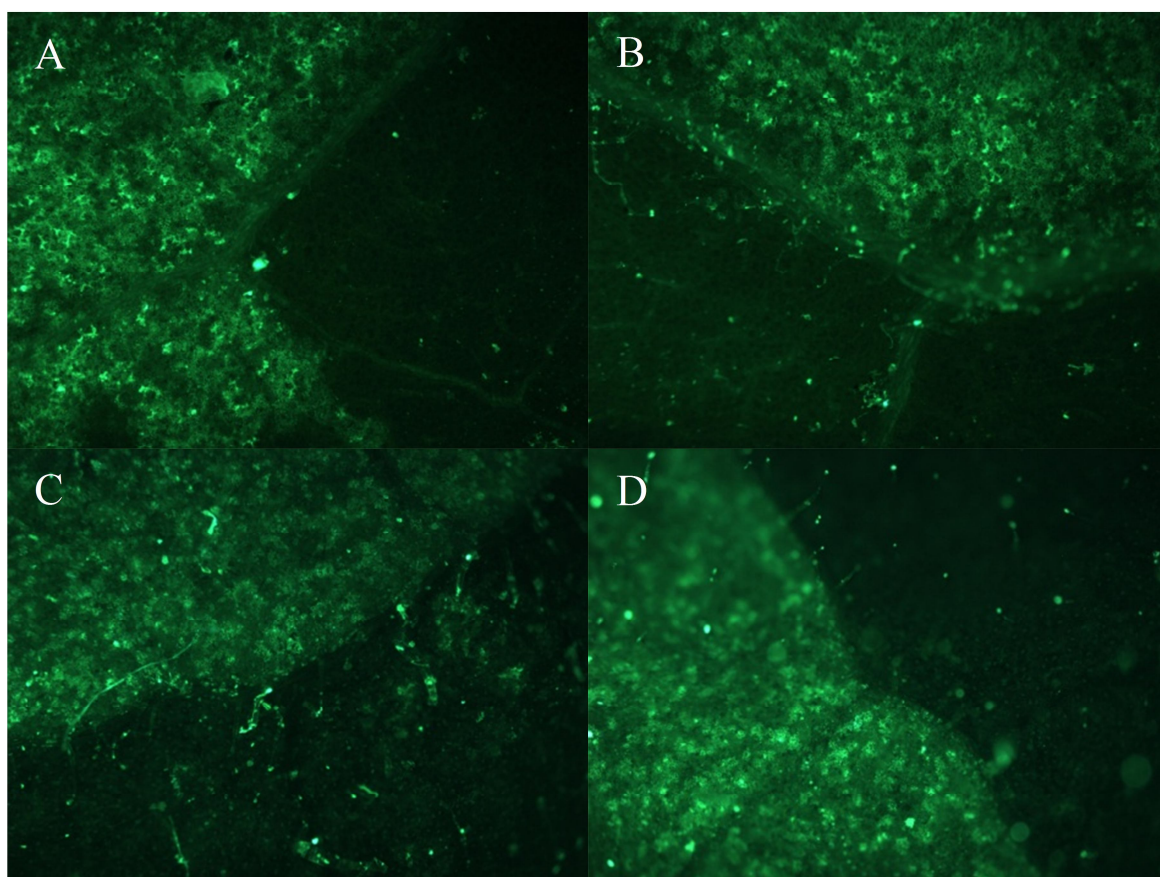


Figure 3.10 GFP expression in Nt-ANT1 transgenic and wild type tobacco leaves infiltrated with pRS13, second infiltration experiment. Photos show boundaries between infiltrated and non-infiltrated leaf areas. The infiltrated areas are evident by the presence of multiple cells fluorescing bright green. The round tips of the glandular trichomes are also visible, fluorescing green, throughout the leaf. A, GH23720 Nt WT. B, GH23725 Nt WT. C, GH23732 Nt-ANT1. D, GH23733 Nt-ANT1.

The results of the third experiment (Section 3.2.9), which tested combinations of P19, ANT1 and pRS13 at different dilutions, showed anthocyanins becoming clearly visible seven days post-infiltration as a speckled red tinge in the ANT1 infiltrations of wild type tobacco plants (Figure 3.11 A). At this point the 1/8 and 1/16 dilutions gave the darkest colour change with 1/32, 1/4 and 1/2 dilutions a bit fainter and 1/1 dilution very faint. At 14 days post-

infiltration the 1/1 infiltration was visible but fainter than the other diluted infiltrations which were similar in colour intensity (Figure 3.11 E).

Comparison of the four different infiltrates showed the effect of P19 on the infiltrations (Figure 3.11). The ANT1 and pRS13 infiltrated areas without P19 remained healthy throughout the life of the leaf (Figure 3.11 A, C, E and G). The ANT1+P19 and pRS13+P19 infiltrated areas both turned yellow at 7 days post-infiltration at all dilutions (Figure 3.11 B) and the pRS13+P19 infiltrations even became necrotic in some areas (Figure 3.11 D). At 14 days post-infiltration this effect was more pronounced. Clearly the presence of *A. tumefaciens* carrying the P19 plasmid was detrimental to the infiltrated leaf areas.

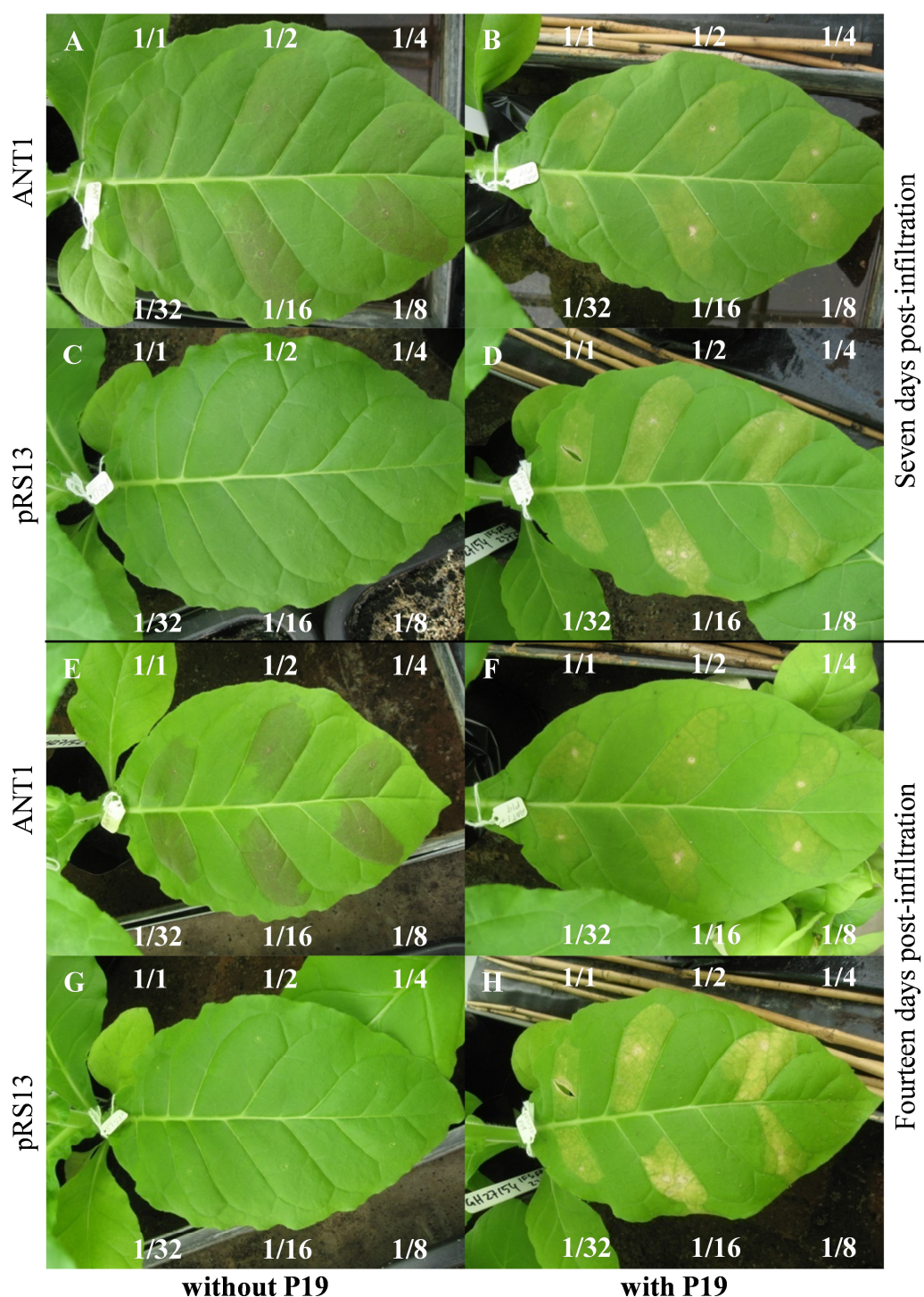


Figure 3.11 Serial dilution of *A. tumefaciens* leaf infiltrates, ANT1 or pRS13 (labelled left), without (left) or with (right) P19, at seven (upper half) or fourteen (lower half) days post-infiltration in wild type tobacco plants. The dilutions of the infiltrate for each of the six infiltrations per leaf are labelled around each leaf.

Since the 1/2 to 1/32 dilutions in the previous experiment gave very similar results, the fourth experiment (Section 3.2.9) tested dilutions down to 1/1024th of the original ANT1 *A. tumefaciens* suspensions in wild type tobacco plants. This infiltration was not as effective as the previous and less red colour developed in the infiltrated areas (Figure 3.12). It was sufficient, however, to determine that the 1/16 and 1/64 dilutions were more effective than the others.

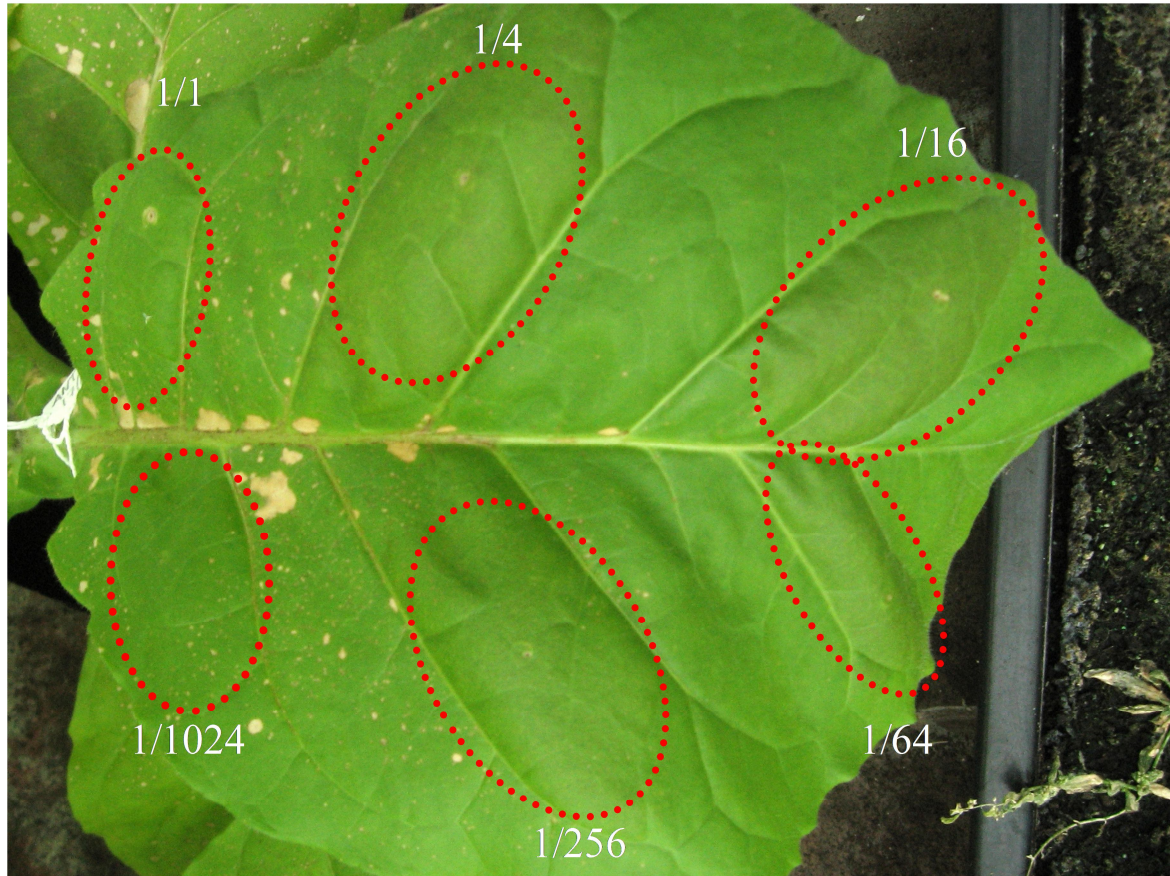


Figure 3.12 Serial dilution of ANT1 *A. tumefaciens* leaf infiltrate in wild-type tobacco plants. ANT1 infiltration areas, 12 days post-infiltration, are circled and the dilution of the infiltrate for each of the six infiltrations is labelled around the leaf. A faint reddish colour is present in the infiltration areas, most noticeable in the 1/16 and 1/64 dilutions. The brown lesions on the leaf were unintentionally caused by pesticide spray damage (chemical unknown).

The fifth experiment was carried out to test whether infiltration of the genistein precursor, naringenin, into the seven day old IFS/CHI infiltration site in Nt-ANT1 transgenic tobacco plants allowed isoflavonoid production at levels detectable by HPLC (Section 3.2.9). After seven days the sites that were only infiltrated with infiltration buffer were physically identical to uninfiltrated areas. The sites infiltrated with IFS/CHI had noticeably reduced pigment levels and were greener than the bronze red colour of the surrounding leaf (Figure 3.13).

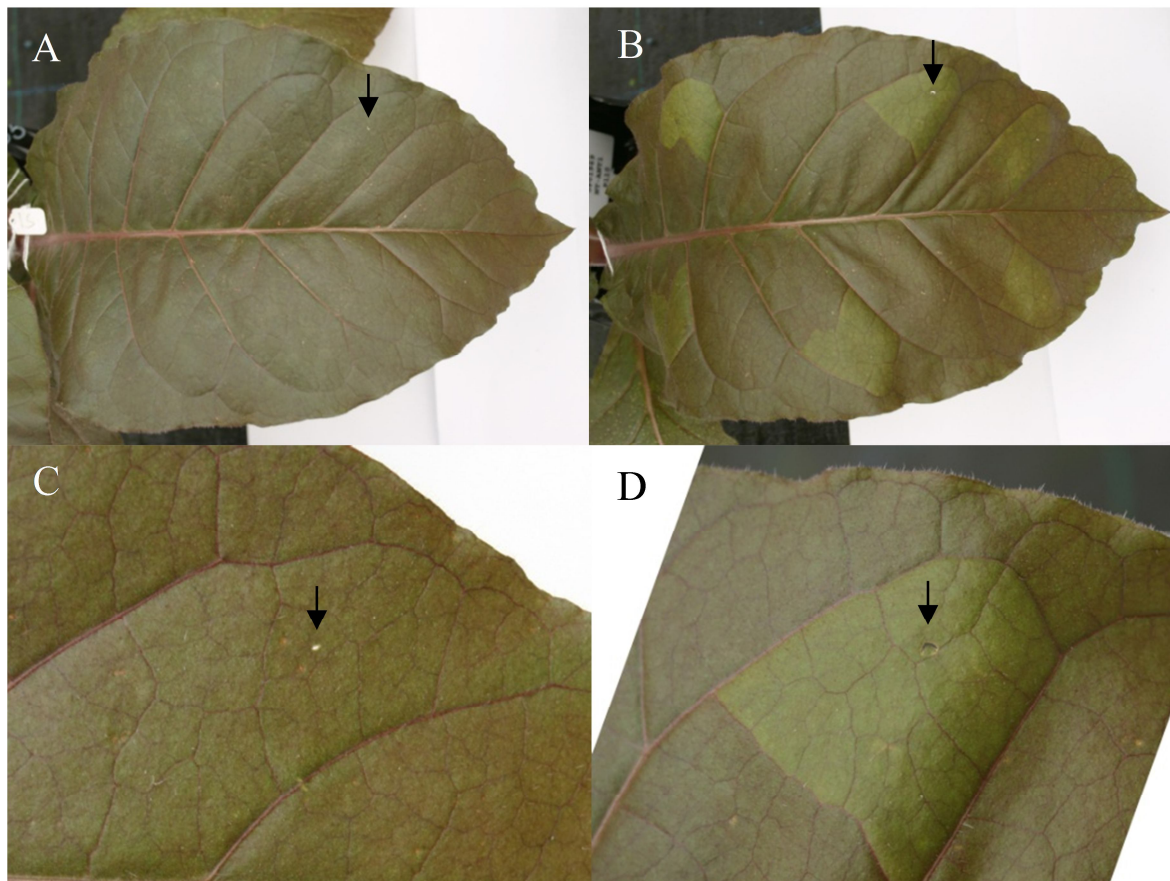


Figure 3.13 Nt-ANT1 leaves infiltrated without or with *A. tumefaciens* carrying IFS/CHI.

A, infiltrated with infiltration solution only. B, infiltrated with IFS/CHI. C, detail of one infiltration site in A. D, detail of one infiltration site from B. The arrows point to the corresponding infiltration site entry holes.

The HPLC chromatogram had very small, if any, peaks at the elution times of isoflavonoids and, since PDA spectra become less clear as peak size decreases, none of the small peaks could be identified as isoflavonoids (data not shown). Any isoflavonoids, if present, were below the detection limit of this technique.

3.3.4 Isoflavonoid content in infiltration experiments determined by LCMS

Leaves of Nt-ANT1 plants were infiltrated with IFS2_12 or IFS/CHI, or Nt-IFS/CHI plants were infiltrated with or without ANT1 in the sixth experiment using the optimised procedures from experiments one to five (Section 3.2.9). All of these samples which had an IFS gene present (as either stable or transient transgene) produced LCMS MS2 peaks of m/z 271 with RDA fragments of m/z 153 (Figure 3.14, see Figure 3.1 for the major isoflavonoid compound and fragment masses). This confirmed that genistein and its conjugates were produced in these tissues.

Genistin (genistein glucoside) was the major genistein component, being present at 2 to 9 times the concentration of genistin malonate (Figure 3.15). The genistein aglycone was present in much lower concentrations. The levels of genistein and genistein conjugates in Nt-ANT1 plants ranged from 0.07 to 1.26 $\mu\text{g g}^{-1}$ DW leaf tissue. Uninfiltrated Nt-ANT1 leaves contained no genistein or genistein conjugates and neither did Nt-ANT1 leaves infiltrated with the control vector pHZbar. The two stably transformed Nt-IFS/CHI plants produced genistein without any infiltrate at lower levels of 0.04 to 0.15 $\mu\text{g g}^{-1}$ DW. When infiltrated with ANT1 the amount of genistein and genistein conjugates increased. In Nt-IFS/CHI plant 5P02034 infiltrated with ANT1, the amount of genistein and genistein conjugates was the highest level detected – a total of 4.52 $\mu\text{g g}^{-1}$ DW and in plant 5P02035 it was 0.21 $\mu\text{g g}^{-1}$ DW. The Nt-IFS/CHI plants produced no genistein or genistein conjugates when infiltrated with pHZbar (Figure 3.15). No genistein or genistein conjugates were detected in a no-sample extraction or in any samples that did not contain an IFS gene – i.e. the control NT-ANT1 plants without infiltration or with the control plasmid infiltrated produced no genistein or genistein conjugates (Figure 3.15).

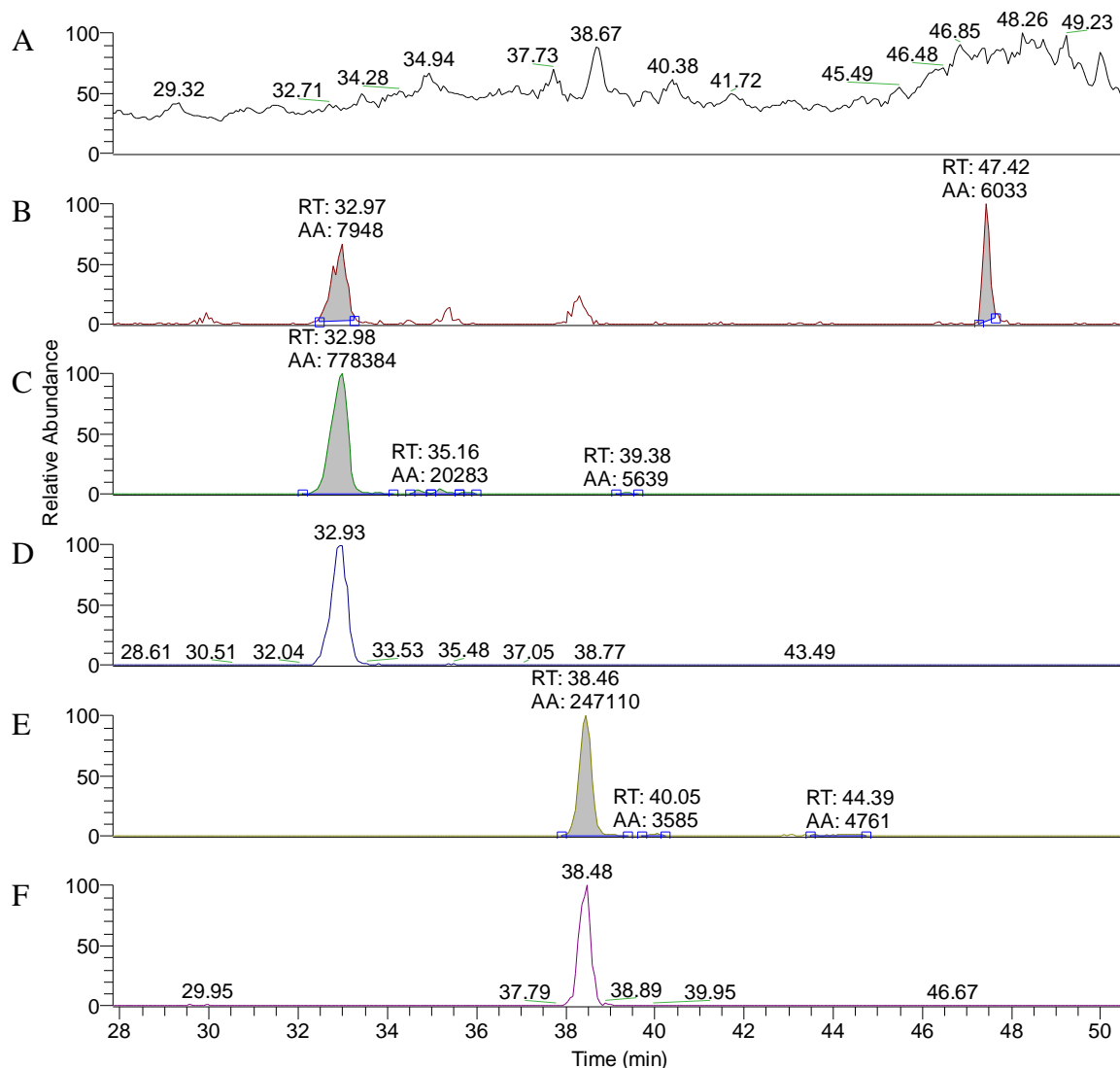


Figure 3.14 LCMS spectra of leaf tissue isoflavonoid extract from Nt-IFS/CHI transformant 5P02034 infiltrated with ANT1. A, MS1 total ion count of all masses over retention times of 28-50 min. B, genistein and genistein conjugates - MS2 fragments of m/z 153 from MS1 ions of m/z 271. C, genistin - MS2 fragments of m/z 271 from MS1 ions of m/z 433. D, genistin - MS3 fragments of m/z 153 derived from the MS2 fragments in C. E, genistin malonate - MS2 fragments of m/z 271 from MS1 ions of m/z 519. F, genistin malonate - MS3 fragments of m/z 153 derived from the MS2 fragments in E.

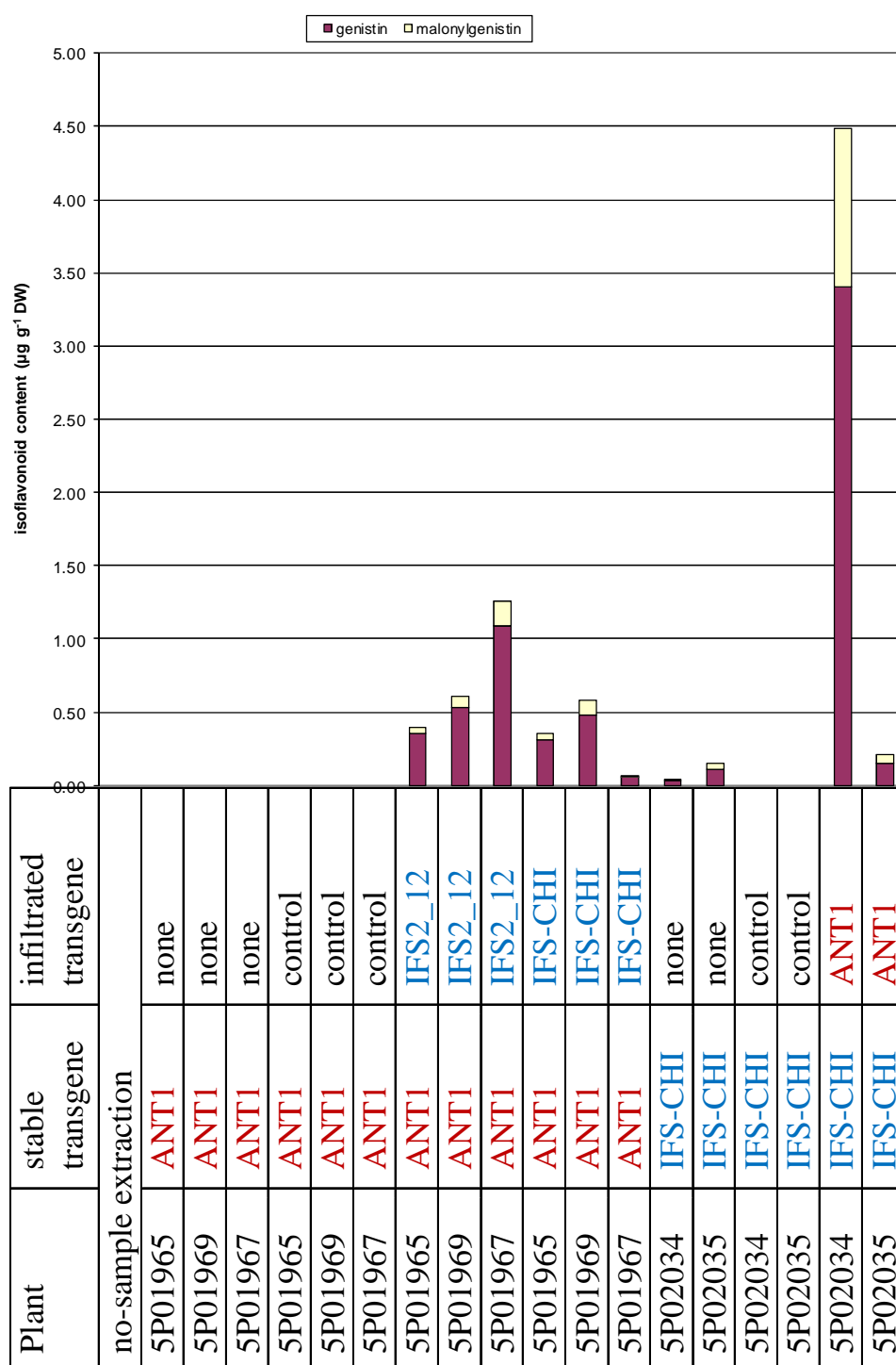


Figure 3.15 Isoflavonoid content of tobacco plants with combinations of stable and infiltrated IFS, ANT1 and negative control (pHZbar) transgenes. Genistein aglycone is not shown because the concentrations that were detected were too low to register on this chart.

3.3.5 Isoflavonoid content in the progeny of stable tobacco transformant crosses

Seedlings were grown from the self-fertilised seed of tobacco transformants Nt-IFS2_12 23722, Nt-ANT1 5P01963, Nt-IFS/CHI plants 5P02034, 5P02035 and 5P02036. Seedlings were also grown from crosses 5P02034x5P01963, 5P02036x5P01963, 5P01963x5P02034 and 5P01963x5P02035, where the first genotype is the maternal plant and the second is the pollen donor. Isoflavonoid extracts from the progeny were analysed by LCMS (Section 3.2.10). RDA fragment analysis again confirmed the presence of genistein in the forms of genistein, genistin and genistin malonate (Figure 3.16).

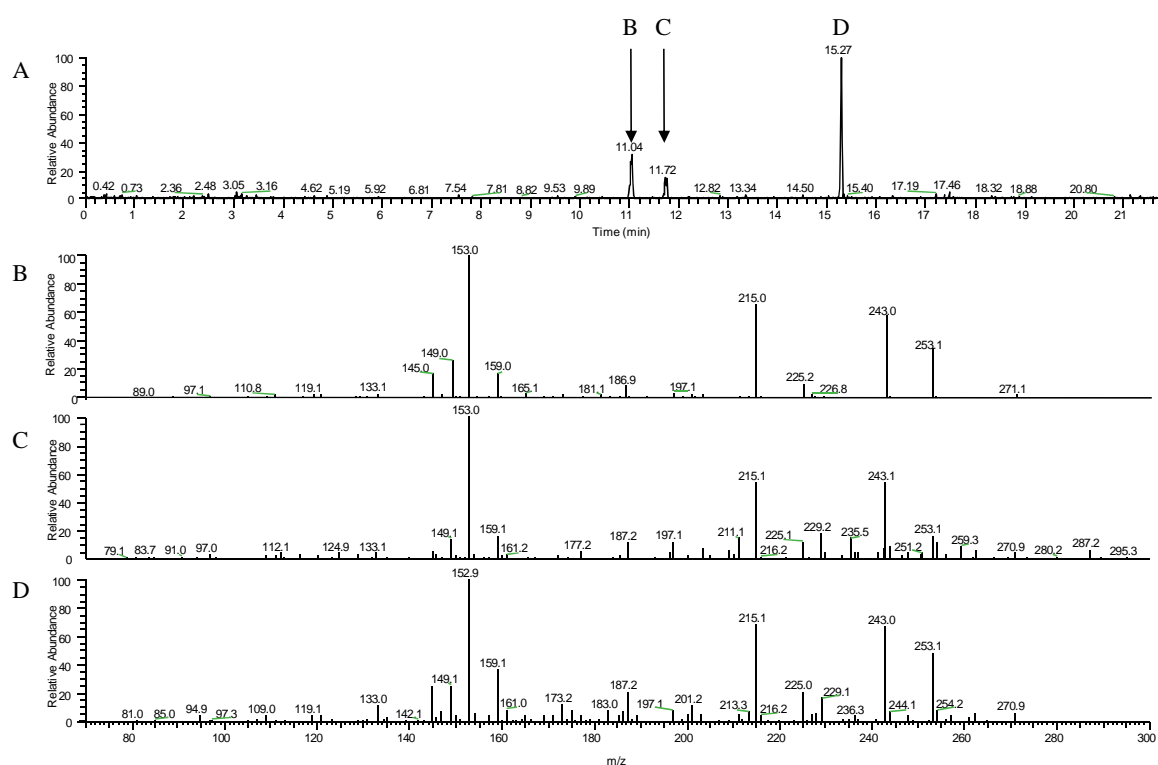


Figure 3.16 Mass spectra of three peaks containing MS2 fragments of m/z 271. A, MS3 fragments of m/z 153 derived from MS2 fragments of m/z 271 over the retention times of 0-21 min. B, genistein - peak in A and MS3 mass spectrum of that peak. C, genistin malonate - peak in A and MS3 mass spectrum of that peak. D, genistein - peak in A and MS3 mass spectrum of that peak.

Quantification of the three compounds again found genistin to be the most common form with genistein being second-most prevalent at concentrations of 1/20th to 1/200th of the amount of genistin. In this experiment, genistin malonate was the lowest concentration at about a quarter of the genistein level. Genistin was 95 to 100% of the total genistein content.

All of the plants that contained no transgene or only ANT1 produced no detectable genistein or genistein conjugates. Of the 30 plants that were progeny of IFS2_12 or IFS/CHI transformants, 13 produced a detectable level of genistin ranging from 0.006 to 19.2 $\mu\text{g g}^{-1}$ DW. Only six plants produced more than 0.1 $\mu\text{g g}^{-1}$ DW genistin although two produced 6.6 and 19.2 $\mu\text{g g}^{-1}$ DW genistin respectively, greater than the 3.4 $\mu\text{g g}^{-1}$ DW genistin which was the highest level found previously, in the ANT1 infiltration of 5P02034. The highest level was detected in a purple plant which was expressing ANT1 but there was no clear correlation between genistin levels and the colour phenotype (Figure 3.17). Progeny from plant 5P02034 (selfed, maternal and paternal) had the highest rate of genistin production with six out of ten plants producing genistin. Three of the nine progeny from plant 5P02035 produced genistin, while only one of the nine progeny from plant 5P02036 produced genistin.

3.4 Discussion

Transformation of tobacco as a model plant was intended to test the hypothesis that this system would be able to show whether cloned white clover isoflavonoid biosynthesis genes were functional and this has been confirmed for IFS2_12.

The leaf infiltration method of the Sainsbury Laboratory (Section 3.2.8) was developed for *N. benthamiana* and, contrary to expectations, the method did not work initially with *N. tabacum*, requiring a process of testing and optimisation until a working protocol was established. Leaf infiltration is a relatively fast process once plants, vectors and *A. tumefaciens* cultures are in place but as it turned out this system still required some method development. Tobacco and *N. benthamiana* plants were transformed with ANT1 to create plants that were amenable to leaf infiltration and were likely to accumulate higher levels of precursors common in the biosynthesis pathways leading to isoflavonoids and anthocyanins, namely naringenin.

Infiltrations were unsuccessful at first. The first infiltration was into young leaves which were severely affected by all infiltrates and the infiltrated areas collapsed, even if just food colour in water was infiltrated (Figure 3.8). Subsequent experiments showed that infiltration buffer had no effect on older leaves (Figure 3.13), so it is most likely that very young leaves are severely affected simply by the infiltration of liquid. The presence of GFP in leaf areas infiltrated with pRS13 indicated that the infiltration did result in transient expression of the infiltrated transgenes. Unfortunately the infiltrated leaf tissues were so badly affected that after five days they were on the verge of death and were not a good model for transgenic plant expression. The leaf probably did not have sufficient time to accumulate isoflavonoids and the wilting and death of the tissue would be complicating factors in the analysis of its metabolic profile. Similarly there was possibly insufficient time to produce or accumulate sufficient β -glucuronidase which is why the leaves were GUS negative. While these infiltrated areas were analysed by HPLC, the production of isoflavonoids would have been too low to detect by HPLC, so no conclusion can be drawn whether these infiltrations produced isoflavonoids.

Due to the result of the first experiment, the second experiment tested the effect of infiltration of water and of infiltration solution without *A. tumefaciens* on more mature leaves. These areas were visually unaffected by the infiltration so the infiltration process itself, and the infiltration buffer, are not harmful to tobacco leaves of sufficient age. The presence of GFP protein in pRS13 infiltrated areas again confirmed that the transient expression system worked but no anthocyanins were visible in the ANT1 infiltrated areas and no isoflavonoids detected, so the system was not able to express these genes sufficiently to make a detectable product.

The third experiment was intended to test the effect of varying concentrations of *A. tumefaciens* on the infiltrated areas. The concern was that the *A. tumefaciens* bacteria themselves were affecting the plant and lower concentrations might reduce this effect while still performing the transformation. To eliminate the complicating factor of mixing two *A. tumefaciens* cultures, individual *A. tumefaciens* cultures with ANT1 or GFP were infiltrated, as well as these cultures mixed with *A. tumefaciens* carrying P19. The result was dramatic and unexpected – the infiltrations with P19 turned yellow while those without P19 remained green and healthy (Figure 3.11). After five days, anthocyanin production became visible as a red colouration of the ANT1 infiltrated areas. The P19 protein had previously been shown to suppress post transcriptional gene silencing and thus increase the production of transiently expressed transgene products and also maintain this production until senescence of the leaf in *N. benthamiana* (Voinnet et al. 2003). This experiment showed that the P19 protein itself or some other factor of the P19 vector was toxic to the leaf tissue and prevented it from expressing higher levels of the co-infiltrated transgene.

Regarding the *A. tumefaciens* dilutions, the original concentration of *A. tumefaciens* – 20 ml of bacterial culture centrifuged and resuspended in 10 ml of infiltration buffer - was found to be the least effective and dilutions of this ranging from 1/2 to 1/32 were almost equally effective in eliciting anthocyanins in the ANT1 infiltrated areas. This indicates that *A. tumefaciens* have already saturated the leaf at 1/32 dilutions in terms of transgene expression which suggests that the optimal level may be at greater dilutions still. A further experiment that tested greater dilutions of up to 1/1024 had a lower ANT1 expression overall than the previous experiment but still showed that the optimal dilution of *A. tumefaciens* lay between 1/4 and 1/256. Given the wide range of *A. tumefaciens* concentrations in successful

infiltrations, a convenient midrange value of 1/10 dilution of the normal *A. tumefaciens* suspension was used for subsequent experiments.

When bronze-coloured ANT1 leaves were infiltrated with IFS/CHI or with infiltration buffer for the fifth experiment, after seven days the IFS/CHI infiltrated areas were less red-coloured than the surrounding uninfiltrated areas or the areas with only infiltration buffer (Figure 3.13). This indicated that a constituent of the infiltrate was reducing the anthocyanin concentration. It is possible that the presence of *A. tumefaciens* had a general effect on the tissues with a side-effect of reducing anthocyanin content. Alternatively, the introduction of IFS/CHI into these cells may have resulted in a depletion of naringenin, the common precursor to isoflavonoids and anthocyanins, and thus caused a reduced production and accumulation of anthocyanins which may, as a result of turnover, actually decrease in concentration. It would be of interest to test the possibility that IFS/CHI is causing the reduction of anthocyanins since this would then be a visual indicator of the biosynthesis of metabolites, including flavonoids, isoflavonoids and condensed tannins that share the anthocyanin precursors. Xie et al. (2006) produced transgenic tobacco plants expressing both PAP-1 and anthocyanidin reductase (ANR) to produce condensed tannins. The plants that expressed PAP-1 and ANR transgenes contained lower levels of anthocyanins than those expressing only PAP1. This is an example of how an anthocyanin-inducing transcription factor co-expressed with an enzyme utilising the same precursors could be used as a general screen for the presence of biosynthesis of some compounds.

The HPLC tests of the IFS/CHI infiltration areas, with or without infiltrated naringin, failed to show any isoflavonoid peaks, and neither did any previous tests of tobacco samples. Given subsequent results discussed below, where all samples with the same ANT1 tobacco and IFS/CHI infiltrate produced genistein, it is likely that these samples also produced low levels of genistein but that HPLC with PDA detection was not sensitive enough to detect genistein in tobacco samples and that a more sensitive detection technique, such as mass spectrometry, was essential for the successful functional testing of IFS genes in tobacco.

In the final infiltration experiment the optimised method was used and some changes were made to increase the likelihood of detecting low concentrations of isoflavonoids. An area approximately 3x the previous infiltration sites (one half of a leaf, about 150 mg DW) was

infiltrated with the diluted *A. tumefaciens* to increase the amount of tissue that could be extracted. This tissue was extracted three times with acetone to ensure the extraction of the maximum amount of isoflavonoids present in the tissue. The samples were analysed by LCMS to allow the detection of very small amounts of isoflavonoids with water blanks in between each sample to minimise any possible carryover of sample in the LCMS. This experiment worked well – the infiltrated areas suffered no adverse effects and the ANT1 infiltrated areas produced visible anthocyanins. LCMS analysis of the extracted tissue was sensitive enough to detect the presence of low levels of genistein conjugates and confirmed their identity by detecting RDA fragments of m/z 153 derived from MS2 fragments of m/z 271 which are unique to genistein. This result occurred in all Nt-ANT1 leaves infiltrated with IFS2_12 or IFS/CHI and in the IFS/CHI transformants that were uninfiltrated or infiltrated with ANT1 (Figure 3.15). This confirms that both IFS2_12 from white clover and IFS/CHI (IFS from soybean) are functional isoflavone synthase genes and supports the hypothesis that the tobacco transient expression system can be used to perform functional testing of isoflavonoid biosynthesis genes.

Control samples of Nt-ANT1 plants with no infiltration or infiltrated with the empty vector pHZbar had no detectable levels of genistein. This confirms that these plants do not naturally produce isoflavonoids and that the infiltration with an empty vector does not elicit any latent production of isoflavonoids either. The no-sample extraction control also contained no detectable genistein which gives confidence that no genistein is being inadvertently introduced or cross-contaminated in the extraction process. This is an important check for a technique that can detect extremely low amounts of compounds – the lowest quantified genistin malonate peak was 1.6 picograms in the 20 μ l sample injection.

The presence of genistein in Nt-ANT1 leaves infiltrated with IFS2_12 or IFS/CHI show that the developed method of transient expression of transgenes through *A. tumefaciens* leaf infiltration is able to express a metabolic biosynthesis enzyme and produce detectable levels of the product of that enzyme 7 days post-infiltration.

The detection of genistein glycosides in uninfiltrated Nt-IFS/CHI leaves shows that these stable transformants are also able to produce isoflavonoids in leaves through the action of the IFS/CHI fusion protein. The CHI component of this enzyme fusion produces the

precursor to genistein, naringenin, so it may have increased the amount of available naringenin in these transformants. Once ANT1 was infiltrated into these leaves the production of genistein increased moderately in one sample and to a larger extent in another.

The fact that the compounds detected in the plants were almost entirely genistin and genistin malonate shows that existing glucosyl and malonyl transferases are active in tobacco and can add glucose and malonate groups to the aglycone genistein. Glycosylated isoflavonoids are more water-soluble than the aglycone and glycosylation is required for their transport into the vacuole (Jones and Vogt, 2001) which may aid the accumulation of genistein.

Given the variability of genistein levels in the samples and the low number of biological replicates it cannot be argued conclusively that one system is more or less productive than another although some trends were observed. The two samples that contained only the IFS/CHI transgene produced lower levels of genistein than seven out of the eight samples that contained either IFS2_12 or IFS/CHI together with ANT1. This supports the previous findings by Tian and Dixon (2006) and confirms the hypothesis that the up-regulation of anthocyanin biosynthesis by expression of the transcription factor ANT1 is also able to increase the production of isoflavonoids in tissues expressing transgenic IFS enzymes. It supports the suggestion that the increased production of naringenin precursor due to expression of ANT1 is able to be used by IFS to produce an increased amount of genistein.

Tests of progeny from crosses of Nt-ANT1 and Nt-IFS/CHI plants showed that the transgenes were inherited and active in progeny and both transgenes could be crossed into one genotype. Although these plants were not PCR tested for the presence of the transgenes, their phenotypes of coloured leaves and genistein production confirmed the presence of ANT1 and IFS/CHI respectively. Genetically, when inherited in a Mendelian fashion, about 50% of crossed progeny and 75% of selfed progeny will inherit at least one copy of the transgene (as long as the transgene has integrated into one location in the genome) so some of the progeny were expected to contain no transgene and be unable to produce genistin. The ratio holds true for plant 5P02034 with 2/3 (66%) of selfed progeny and 4/7 (57%) of crossed progeny producing genistin. It does not match the other plants though since all of the IFS2_12 23722 progeny, 33% of the 5P02035 progeny and 11% of the 5P02036 progeny produced genistin. It is possible that some of these plants inherited the transgene but that it

was not expressed or expressed at a very low level so that no genistin was produced. Epigenetic effects can affect transgene expression depending on copy number and where in the genome it integrates (van der Krol et al., 1990) and these effects would be different in every primary transformant and inherited by the progeny. 5P02036 was not tested previously and it appears to contain a transgene that has low activity, resulting in progeny where the IFS/CHI gene is present but genistein production is very low or turned off. Conversely 5P02034 produced the highest genistein concentration in the last infiltration experiment and its inheritance pattern suggests that IFS/CHI is active in all of the progeny that contain the gene.

The confirmation of the functionality of IFS2_12 and IFS/CHI makes it possible to confidently use these genes to test the hypothesis that overexpression of IFS leads to higher levels of isoflavonoids in transgenic white clover plants. This is tested in the following chapter.

4 Transgenic white clover

4.1 Introduction

The primary goal of this thesis was to test if isoflavonoid levels in white clover could be increased through metabolic engineering. The hypothesis tested was that increased IFS expression in white clover would increase the concentration of isoflavonoids. To test the hypothesis, white clover plants were transformed to overexpress IFS, using both the endogenous IFS2_12 (whose isolation was described in Chapter 2) and an IFS/CHI fusion gene derived from soybean and alfalfa (Tian and Dixon, 2006), both of which were confirmed to be functional in tobacco (Chapter 3).

The standard white clover transformation system, used to overexpress IFS2_12 in white clover (var. Huia), uses genetically diverse seeds as explants and, consequently, each genotype could have qualitatively and quantitatively different isoflavonoids. This background variability makes it less likely that moderate changes in isoflavonoid levels, caused by the transgene, can be distinguished as being different from the wild-type. For the subsequent overexpression of the IFS/CHI fusion gene in white clover, a novel method was developed where the two cotyledons of a seed were tracked and maintained separately as clones in tissue culture, with one cotyledon being transformed and the other remaining as wild type. This use of a wild type control clone allows the determination of the background isoflavonoid level for each genotype and this can be subtracted from the level found in the transgenic plant to isolate the effect of the transgene on isoflavonoid content.

4.2 Materials and Methods

4.2.1 Plant materials

White clover (*Trifolium repens*) cultivar 'Huia' (Margot Forde Germplasm Centre accession number C19739) was the cultivar used for IFS and IFS/CHI transformations.

Other cultivars and individuals used for isoflavonoid analysis included: Grasslands ‘Pawera’ cultivar, accession number FT1179, an estrogenic (high isoflavonoid) red clover (*T. pratense*).

Grasslands ‘G27’ cultivar, accession number FT1184, a non-estrogenic (low isoflavonoid) red clover bred from ‘Pawera’.

Grasslands ‘Colenso’ cultivar, accession number F3309, a medium level estrogenic (medium formononetin) red clover.

Grasslands ‘Sensation’ cultivar, accession number F2903, red clover.

Grasslands ‘Crossway’ cultivar, accession number F3092, a creeping cultivar of red clover.

White clover cultivar ‘Sustain’.

White clover Grasslands Mapping Population 1 parent 1, genotype ‘21125’.

White clover Grasslands Mapping Population 1 parent 2, genotype ‘20161/21’.

4.2.2 Determining natural isoflavonoid content and IFS gene expression in white and red clover

Ten clones from five non-transgenic plant genotypes from the cultivars ‘Pawera’ red clover, ‘G27’ red clover and ‘Sustain’ white clover and clones from white clover mapping population 1 parents 21125 and 20161/21 were grown under prevailing environmental conditions (Palmerston North, New Zealand, spring). A sample (1 g FW) of leaves taken from each plant was frozen in liquid nitrogen, ground in liquid nitrogen to a fine powder in a mortar and pestle and then freeze-dried. Isoflavonoids were extracted (Section 3.2.12) and separated via HPLC (Section 3.2.13). Samples were compared by plotting peak retention time, PDA max plot peak area and the maximum absorbance wavelength of the PDA spectrum (Figure 4.2). Biochanin A and formononetin compound peaks were also quantified by calculating a linear standard curve from the quantity standard peak areas and converting sample peak areas to micrograms of isoflavonoid aglycone per gram of dry weight tissue.

Leaf tissue was sampled from three white clover plants ('Sustain' plant 1, 21125 plant 1 and 20161/21 plant 1) and three red clover plants ('G27' plant 1, 'Pawera' plant 2 and 'Pawera' plant 7) and ground with a mortar and pestle in liquid nitrogen. The frozen powder was partitioned into a 500 µl volume of sample for RNA extraction (Section 4.2.3) and the remainder was freeze-dried for isoflavonoid extraction (Section 3.2.12).

Leaf, petiole and root tissue was sampled from four cultivars of red clover ('Colenso', 'Crossway', 'Pawera' and 'Sensation') grown under prevailing environmental conditions under shade cloth (Palmerston North, New Zealand, winter). The tissue was extracted (Section 3.2.12) and tested for isoflavonoids as described in Section 3.2.13.

4.2.3 Extraction and quantification of RNA

Leaflets, excluding petioles, were picked from the plants by hand and frozen in liquid nitrogen. Frozen samples were ground by hand using a mortar and pestle. Alternatively the samples were placed in stainless steel grinding jars with a 19 mm stainless steel ball, all cooled in liquid nitrogen, and placed in a ball mill (Retsch MM301) and shaken at 30 Hz for 30 s. The finely ground sample was partitioned into two samples after grinding - a volume of 500 µl of loosely packed powder was used for RNA extraction and the remainder used for isoflavonoid extraction.

RNA was extracted by mixing 1 ml of TRIzol reagent (Invitrogen) with the frozen, ground sample in a 2 ml microtube and incubating it at room temperature for 5 min. 200 µl of chloroform was added, mixed and the sample was incubated at room temperature for 3 min. The sample was centrifuged at 22000 g for 15 min. The supernatant was transferred to a new tube and the remainder discarded. 250 µl isopropanol and 250 µl 0.8 M sodium citrate, 1.2 M sodium chloride were added to the supernatant and the tube was mixed and incubated at room temperature for 10 min. The sample was centrifuged at 22000 g for 15 min at 4°C. The supernatant was discarded and the pellet was washed with 500 µl 70% v/v ethanol and centrifuged at 22000 g for 10 min at 4°C. The supernatant was discarded and the pellet was air dried for 10 min. The RNA pellet was dissolved in 50 µl of RNase-free water.

RNA was quantified using a spectrophotometer (Nanodrop ND-1000). The Tr-IFS2_12 RNA extracts contained between 300 and 1400 ng μl^{-1} RNA (average 825 ng μl^{-1}) with one low sample measuring 60 ng μl^{-1} . The 260/280 nm absorbance ratio was between 1.99 and 2.16 for all samples. To remove genomic DNA contamination, 20 μg of RNA was treated with 20 U DNase1 (Invitrogen) in 30 μl of 5 mM MgSO_4 . The mixture was incubated at 37°C for 1 h, then 75°C for 5 min. To check for any residual DNA contamination, 200 ng of DNase1-treated RNA was used as the template in a PCR reaction with IFS_RT_F1 and IFS_RT_R1 primers (Appendix 4). Since Taq polymerase can only use DNA as template, IFS mRNA will not be amplified. A 35 cycle standard PCR reaction was performed and products were resolved by agarose gel electrophoresis (Section 2.2.8). The positive control produced a band of the predicted size (124 bp) while none of the RNA samples produced any bands (data not shown).

4.2.4 Quantitative PCR of IFS cDNA

IFS transcript levels were measured by quantitative PCR (qPCR) of cDNA from each plant sample. One microgram of RNA was used as template for reverse transcription into cDNA using ThermoScript reverse transcriptase (Invitrogen) and oligo dT primers according to the instructions of the manufacturer. The cDNA was diluted tenfold to serve as a template for qPCR.

To create a standard curve of known template starting quantity for qPCR, the concentration of a pCR2.1+IFSc05 vector positive control plasmid DNA sample was quantified using a Qubit fluorometer (Invitrogen) and the concentration required for 2×10^8 molecules per 5 μl sample was calculated. To calculate the number of molecules in the sample, the following formula was used: $n = (\text{ng} \times 10^{-9} / \text{MW}) \times 6.022 \times 10^{23}$ where n is the number of molecules, ng is the total amount in nanograms and MW is the molecular weight of the molecule. The vector pCR2.1+IFSc05 had a molecular weight of 1668542, so a concentration of 0.1108 ng μl^{-1} was calculated to contain 2×10^8 molecules in a 5 μl sample. A 10x serial dilution, producing samples with 2×10^8 to 2×10^1 molecules per 5 μl , was made and these samples were used as templates for the qPCR standard curve.

The known starting quantity of the standard curve samples was plotted against their threshold cycle for each experiment. The formula of the standard curve trendline was used to calculate the starting quantity (number of transcripts) of the experimental samples from their experimentally determined threshold cycle. In the qPCR reaction, the 2×10^7 to 2×10^2 samples produced an exponential trendline when starting quantity was plotted against threshold cycle. This is seen as a straight line in Figure 4.1 where the starting quantity is plotted on a logarithmic scale. This confirmed that the PCR reagents and protocol functioned as expected.

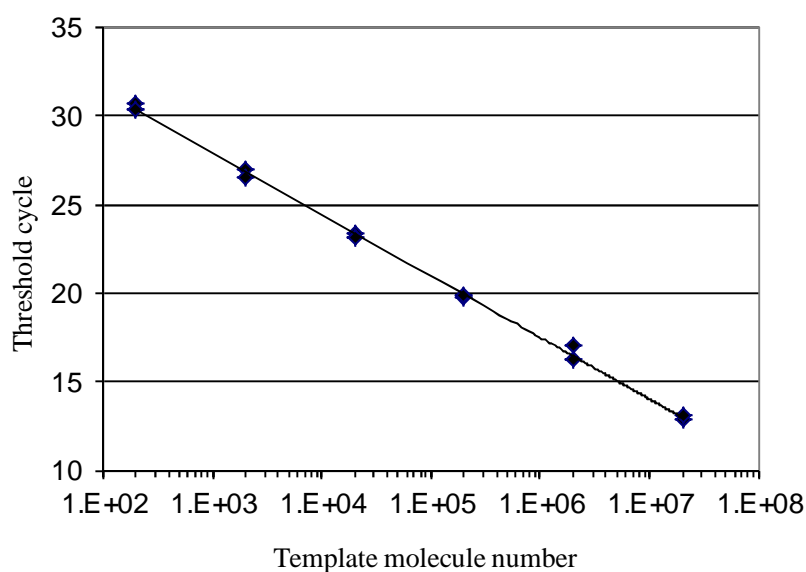


Figure 4.1 IFS qPCR standard curve and trendline

The standard curve trendline had a formula of $y = 10^{11} e^{-0.6614x}$ where y = starting quantity / transcript number and x = threshold cycle. This formula was used to calculate the transcript number of the experimental samples from their threshold cycle value.

qPCR primers were designed as follows. Introns in the genomic DNA sequence are spliced out of the mRNA after transcription and are thus not present in the cDNA. This property was used to design primers that do not anneal to genomic DNA. The qPCR primers were designed to anneal to the cDNA sequence that spanned the intron. Thus only half (about 10 bp) of the primer matches any contiguous part of the genomic DNA. The annealing temperature of a 10 bp sequence is lower than that of the PCR annealing temperature and so the primer is unable to bind to any contaminating genomic DNA during the PCR reaction. The primers used for qPCR were IFS_RT_F1 and IFS_RT_R1 (Appendix 4). These primers match all of the sequence variants cloned from white clover (Chapter 2).

The 20 µl qPCR reactions contained the same concentrations of PCR reagents as in Standard PCR (Section 2.2.5) as well as 10 nM fluorescein and a 0.1333x concentration of SYBR green. Fluorescein is a background fluorescence control for the thermocycler while the SYBR green fluoresces in the presence of double stranded DNA and so allows the direct measurement of the quantity of double stranded DNA PCR product in the well. The sample template was 5 µl of the diluted cDNA, equivalent to 0.5 µl of the reverse transcriptase reaction or 25 ng of RNA template. As no white clover housekeeping gene primers were available at the time of this experiment, the samples were normalised on the basis of total RNA concentration.

Reactions were performed in a MyiQ iCycler (BioRad). The cycling conditions were 95°C for 3 min, then 45 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s. Fluorescence data were collected in every cycle at the annealing step. At the completion of the thermocycling, a melt curve was measured by first denaturing the PCR products at 95°C for 1 min, then the temperature was held at 95°C for 10 s and, for 85 cycles, the temperature was decreased by 0.5°C and held for 10 s per cycle. A fluorescence measurement was made at each temperature setpoint.

The time point at which the SYBR green fluorescence exceeds a set value is the threshold cycle. This value is interpolated to two decimal places from readings around that point. A standard curve plotting threshold cycle against the known starting quantity of the standards allows the calculation of the starting quantity of the experimental samples. The melt curve shows if there is one, or more than one, PCR product present at the end of the reaction by how many peaks are present in the integral of the melt curve. It also shows if the PCR products are different by the location of the peak.

4.2.5 White clover transformation

The protocols of Voisey et al. (1994) and White and Voisey (1994) were amended to transform and regenerate white clover. White clover was transformed with IFS2_12 or IFS/CHI genes in the pRSh1 binary vector (Section 3.2.2). The transgene coding sequence was located between the cauliflower mosaic virus (CaMV) 35S promoter and the octopine synthase (OCS) terminator. The *A. tumefaciens* 'GV3101' strain carrying the desired vector was grown from a frozen glycerol stock inoculation, overnight in 20 ml liquid MGL medium (Appendix 1) containing 100 µg ml⁻¹ spectinomycin, at 28°C with shaking at 200 rpm.

White clover seeds were surface-sterilised by washing with 10 ml 95% v/v ethanol for 30 s, then 10 ml concentrated bleach for 15 min. Seeds were rinsed with sterile water four times then left in 10 ml sterile water on a rotating carousel for 30 min to wash germination inhibiting factors out of the seed coats. Seeds were rinsed twice more with sterile water and 5 ml of water was left in the tube. Sterilised seeds were left at 15°C overnight to imbibe.

Under sterile conditions in a laminar flow hood, the imbibed seeds were dissected under a stereomicroscope (Olympus SZ40) with fine tweezers and a scalpel. The seed coat and endosperm were cut and removed from the seed and the root was cut off. The scalpel was then inserted between the two cotyledons and the seed was sliced into two pieces, exposing the shoot apical meristem. The dissected cotyledons were then placed onto sterile filter paper on CR7 agar media containing no antibiotics or herbicide (Appendix 1).

The *A. tumefaciens* in the overnight culture were pelleted by centrifugation (5 min at 3220 g) and resuspended in 5 ml of sterile 10 mM MgSO₄. All cotyledons to be transformed were

exposed to *A. tumefaciens* by dispensing drops of the suspension onto the dissected cotyledons. Approximately 1000 white clover seeds (2000 cotyledons) of *Trifolium repens* cv 'Huia' were dissected for the IFS2_12 transformations.

For the IFS/CHI transformation, a novel modification of the standard transformation technique was used where the two cotyledons from one seed were tracked and one was left as wild type while the other was transformed. This was named the 'Twinned cotyledon transformation system'. To track cotyledons of the same seed, a grid was drawn onto the filter paper on the non-selective CR7 agar plate with pencil prior to sterilising and the two cotyledons from one seed were placed in matching positions in two labelled plates. Throughout the tissue culture regeneration process, the cotyledons were placed in corresponding positions on two plates which were aligned by a mark on the plates. When one of a pair of cotyledons died or failed to regenerate, both were removed. Wild type plants were grown on plates that did not contain the AMM herbicide selection but did contain antibiotics.

1000 white clover 'Huia' seeds were dissected for the IFS/CHI transformations.

4.2.6 Regeneration of transformed meristem through tissue culture

Plates of *A. tumefaciens*-inoculated dissected cotyledons were sealed with plastic film and were kept at 25°C under fluorescent lighting for 16 h d⁻¹ for 3 d during which the cotyledons expanded and changed colour from white or pale yellow to green. During this time the *A. tumefaciens* transformed the white clover cells along the cut surfaces including the shoot apical meristem. After 3 d, tweezers were used to carefully transfer the cotyledons directly onto CR7 agar medium containing 5 mg l⁻¹ AMM herbicide and 150 mg l⁻¹ TIM antibiotic. The cotyledons were left to initiate and grow shoots for two weeks. Vigorously growing regenerants were then transferred onto selective CR5 plates, containing AMM and TIM, to induce rooting. They were left to grow for two weeks, and some started producing roots at this stage. Regenerants that were green but grew slowly were unlikely to start growing vigorously enough to produce reasonably sized shoots and develop roots and were thus discarded.

Regenerants were then transferred to selective CR0 plates containing AMM and TIM and no hormones. A lack of hormones is root-inducing and the regenerants were left on this medium until they produced roots. Over the following weeks, plantlets with roots were transplanted into sterile peat plugs. The Tr-IFS2_12 plants were named according to the order in which they were transplanted. Up to three regenerated plantlets (A, B and C) from each cotyledon were transplanted – these have either been derived from the same transformation event, from different transformation events or were untransformed tissue that survived the selection process. So the label ‘P01A’ represents plant number 1, replicate A. The Tr-IFS/CHI plants were named similarly to the Tr-IFS2_12 plants but the meaning of the last letter was different - ‘A’ designated wild type plants and ‘B’ for transgenic plants. For replicates, ‘c2’ (copy number 2) or ‘c3’ were added to the end of their names.

Plantlets were left in the peat plugs for one to three weeks until roots emerged at the sides of the peat plugs. The plants were then fully submerged in a chlorothalonil fungicide solution (Bravo, Yates, Auckland, New Zealand) and placed in a covered seedling tray in the containment greenhouse. Over the course of a week, the vents in the lid of the tray were progressively opened and the lid removed to acclimatise the plants to the less humid environment. Then the mesh of the peat plugs was removed and the plants were transplanted into potting mix in PB3 planter bags. Subsequently, the plants remained in the containment greenhouse. When experiments were concluded, plants were maintained in the greenhouse or destroyed by autoclaving.

4.2.7 PCR confirmation of transformation

Genomic DNA was extracted, using a Qiagen Plant mini kit, from leaflets of the 22 Tr-IFS2_12 plants and tested for the transgene insert using the standard PCR protocol (Section 2.2.5) with IFS_F_TrEST and P-3_OCS_rev primers (Appendix 4). PCR products were resolved on an agarose gel (Section 2.2.8) and a band of the predicted size (2041 bp) indicated the presence of the transgene in the plant genomic DNA.

Genomic DNA was extracted, using a Qiagen Plant mini kit, from the Tr-IFS/CHI plants and PCR testing was carried out using pRSH1_F2 and IFS_5'_int_R primers (Appendix 4). A

band of the predicted size (144 bp) indicated the presence of the transgene in the plant genomic DNA.

4.2.8 Determination of isoflavonoid levels in transgenic white clover

Leaf and root tissue from sixteen Tr-IFS2_12 regenerants was collected, freeze-dried and ground in a ball mill as described in Section (Section 3.2.12), and isoflavonoids were extracted and separated by HPLC (Section 3.2.13). The quantity of formononetin glucoside malonate was calculated by comparison of the 254 nm absorbance peak area with the peak area of known standards and correcting for the dry weight of leaf tissue extracted.

Leaf tissue from six Tr-IFS/CHI regenerants and their wild-type clones was collected, freeze-dried and ground in a ball mill as in Section 3.2.12 and isoflavonoids were extracted and separated by HPLC (Section 3.2.13). Formononetin glucoside malonate was calculated as described above.

4.3 Results

4.3.1 Natural isoflavonoid levels and variability in white and red clover

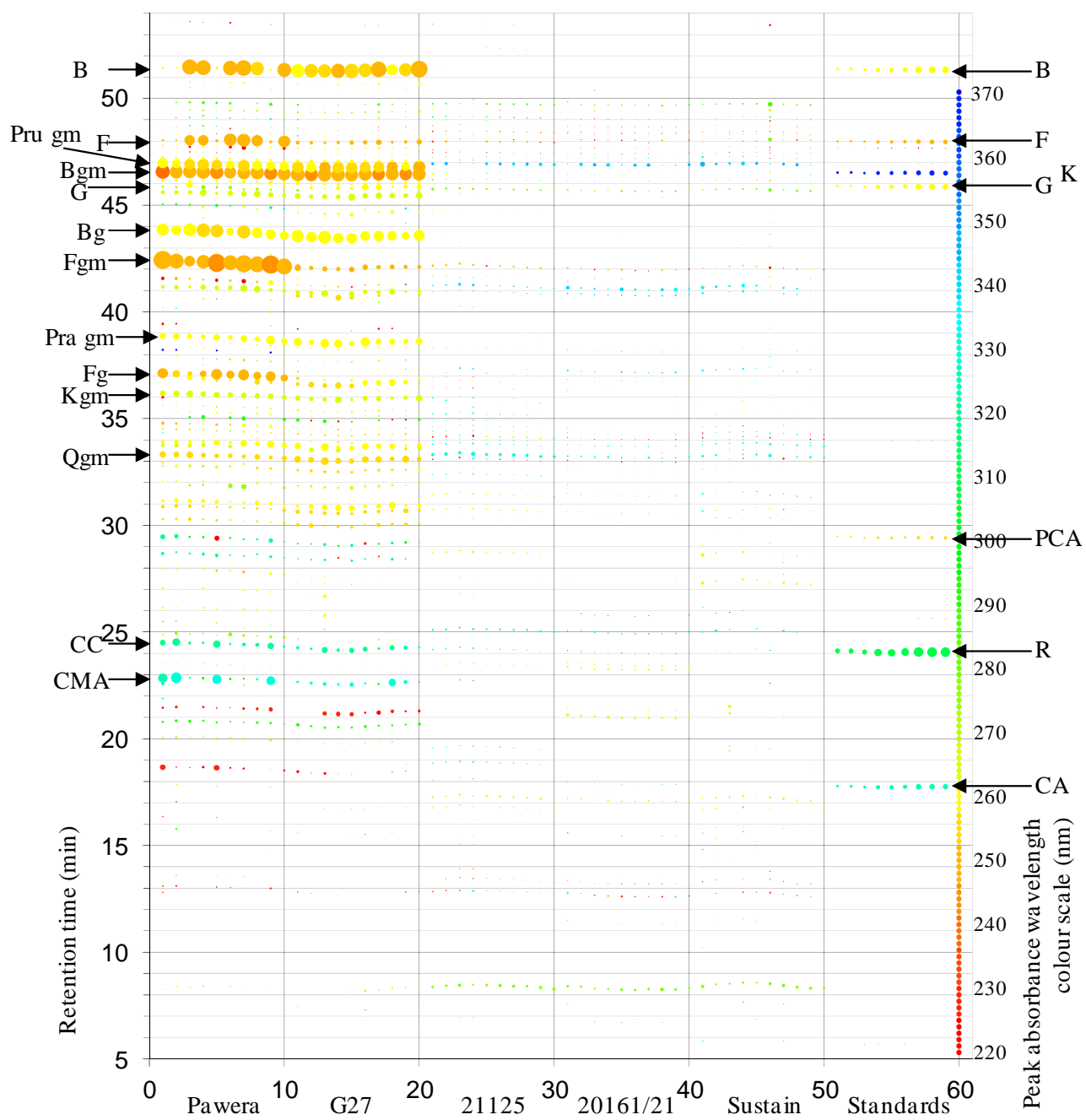
Testing of ten plants each of ‘Pawera’ red clover, ‘G27’ red clover, ‘Sustain’ white clover and clones of white clover 21125 and 20161/21 established that red clover had many times the level of isoflavonoids and other flavonoid compounds than white clover (Figure 4.2), while white clover had such low levels of flavonoids that many of them were not detected. Each cultivar had characteristic patterns and relative quantities of metabolites with a degree of variation within each cultivar. ‘Pawera’ had high levels of formononetin and biochanin A as glucosides and glucoside malonates. There were two chemotypes in ‘Pawera’: four of the ten plants had chemotype A - very low levels of free formononetin (5% of the level of chemotype B) and free biochanin A (3% of the level of chemotype B) with higher levels of cis-clovamide and caffeoyl-malic acid. The other six plants had chemotype B - high levels of the free isoflavonoids and lower levels of cis-clovamide and caffeoyl-malic acid. Chemotype A plants also had higher levels of glycosylated formononetin than chemotype B (48% more) while glycosylated biochanin A was much the same (Figure 4.2).

Red clover 'G27' had broadly the same quantities of the major flavonoids and isoflavonoids as 'Pawera' (chemotype B) – the average level of combined biochanin A compounds was within 10% of the level in 'Pawera'. The exception was for formononetin and its conjugates, where the 'G27' levels were only 9% of the 'Pawera' levels (Figure 4.3).

The three white clover lines had low levels of isoflavonoids, with only formononetin at the retention time of 48 min and formononetin glucoside malonate at the retention time of 42 min, being clearly identifiable in the HPLC results (Figure 4.2). Of these, formononetin glucoside malonate is the isoflavonoid with the highest concentration. The average levels of these two formononetin compounds combined, in 21125, 20161/21 and 'Sustain' are only 1 to 1.6% of the levels found in red clover 'Pawera' (Figure 4.3). Other isoflavonoids, if present, were too low to identify with confidence.

The isoflavonoid content in leaf, petiole and roots of four red clover cultivars is shown in Figure 4.4. All four cultivars had high levels of isoflavonoids ranging from 2500 to 30,000 $\mu\text{g g}^{-1}$ DW (0.25 to 3% DW). The pattern of relative formononetin content in different tissues was variable – 'Pawera' and 'Sensation' had similar levels in the different tissues while 'Colenso' had less in petioles than other tissues and 'Crossway' had much less in leaves and more in roots. The pattern of biochanin A was the same for all cultivars – roughly equal levels in leaves and roots and reduced levels in petioles.

Figure 4.2 ‘Virtual TLC’ (thin layer chromatograph) visualisation of 50 HPLC chromatograms consisting of 10 replicates of five clover cultivars or genotypes. Spot area is proportional to peak area, the Y axis position is the retention time and the colour represents the maximum absorbance wavelength which is dependent on the type of compound. Standards, three replicates each of 25, 50 and 100 $\mu\text{g ml}^{-1}$ of each compound B, F, K, G, PCA, R and CA. Labelled compounds: B, biochanin A; F, formononetin; Pru gm, prunetin glucoside malonate; Bgm, biochanin A glucoside malonate; K, kaempferol; G, genistein; Bg, biochanin A glucoside; Fgm, formononetin glucoside malonate; Pra gm, pratensein glucoside malonate; Fg, formononetin glucoside; Kgm, kaempferol glucoside malonate; Qgm, quercetin glucoside malonate; PCA, p-coumaric acid; CC, cis-clovamide; R, rutin; CMA, caffeoyl-malic acid; CA, caffeic acid;



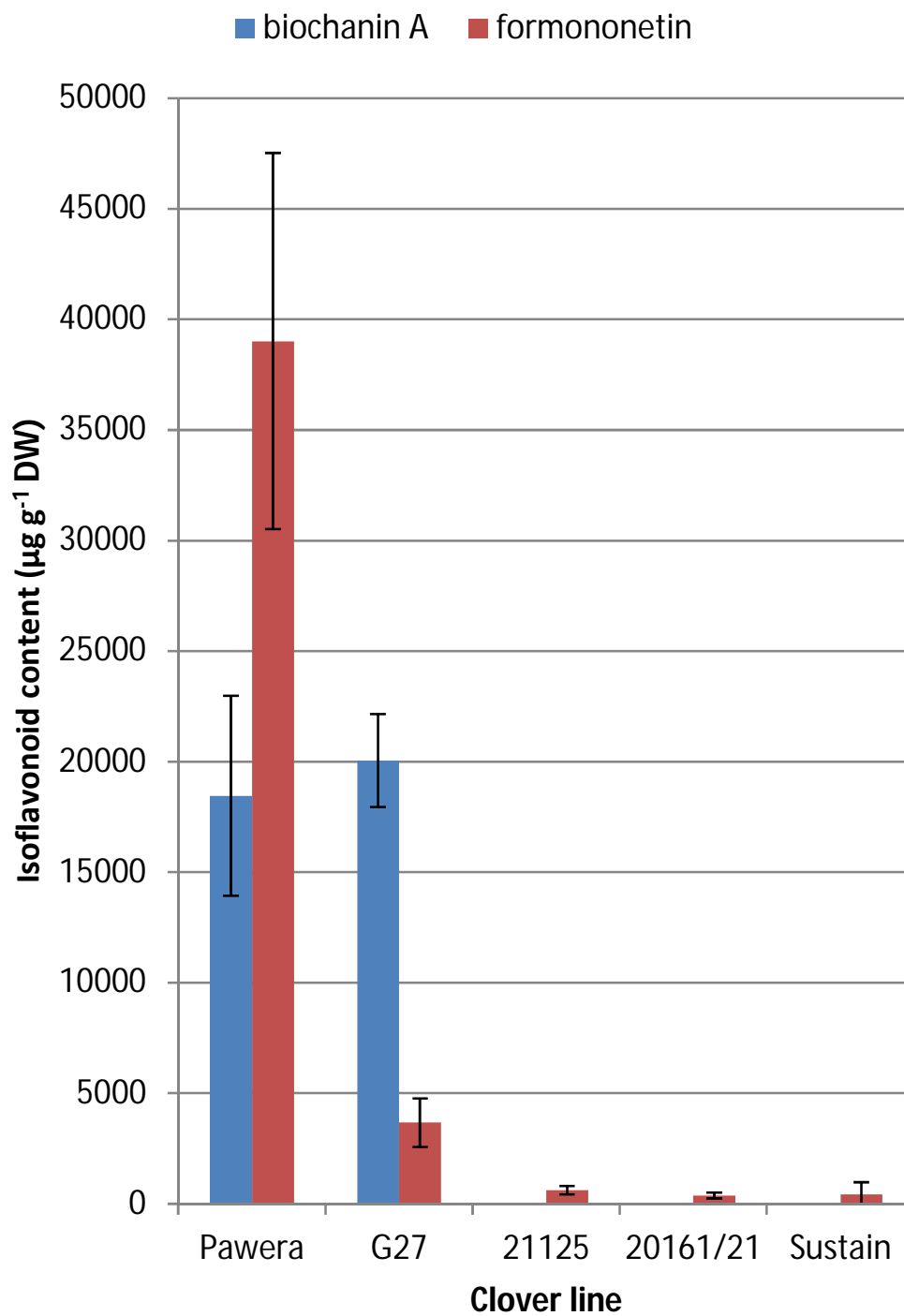


Figure 4.3 Average natural isoflavonoid content of red clover cultivars ‘Pawera’ and ‘G27’ and white clover cultivar ‘Sustain’ and clones of genotypes 21125 and 20161/21. Error bars of one standard deviation are shown.

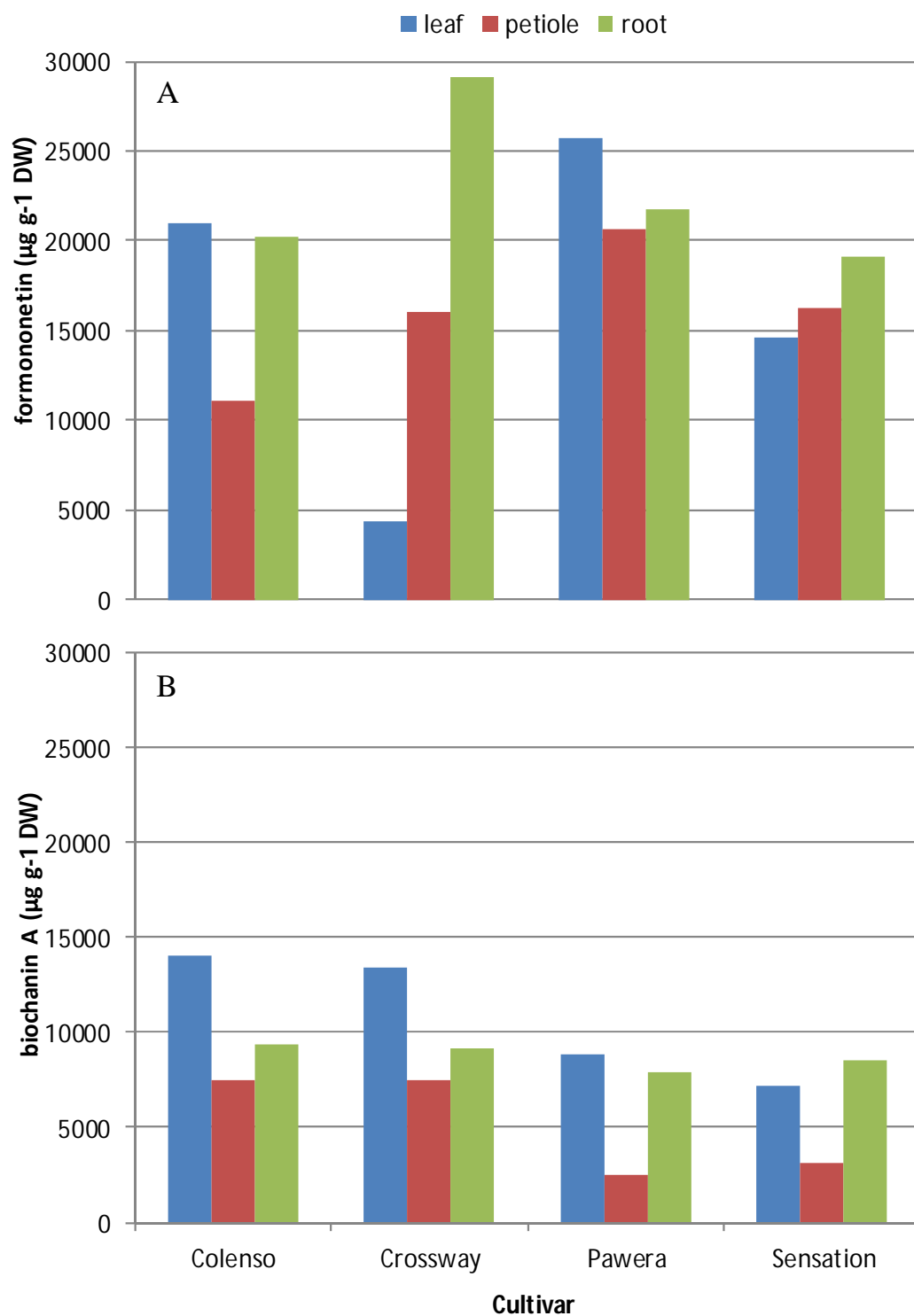


Figure 4.4 Formononetin (A) and biochanin A (B) levels of four red clover cultivars in leaf, petiole and root tissues.

4.3.2 Natural isoflavonoid content and IFS gene expression in white and red clover

The isoflavonoid profile was determined from the same leaf sample used for IFS gene expression and was similar to that of the previous experiment (Figure 4.3) – high levels of formononetin and biochanin A in ‘Pawera’, reduced formononetin in ‘G27’ and low levels of formononetin in the white clover samples (Figure 4.5). Biochanin A was detected in small amounts in the plants 21125 and 20161/21.

In white clover, the IFS gene expression was positively correlated ($R^2=0.999$) with isoflavonoid levels but not in red clover which had IFS levels far lower than those of white clover (Figure 4.5).

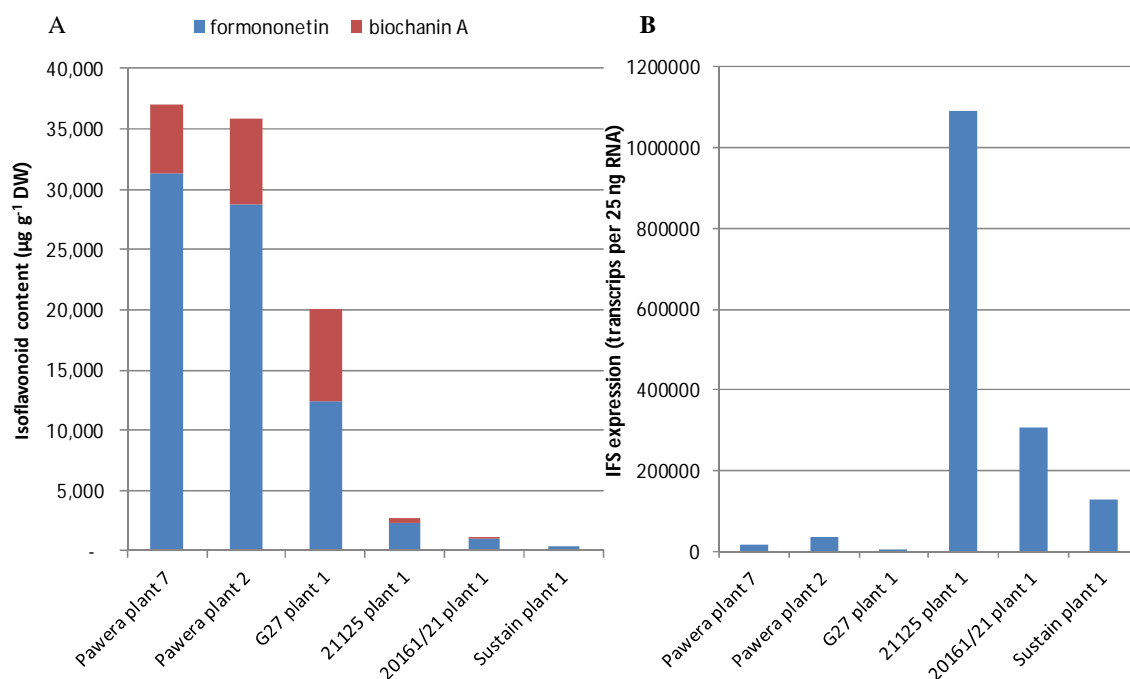


Figure 4.5 Isoflavonoid content (A) and IFS gene expression (B) in wild type red and white clover plants.

4.3.3 Transformation and regeneration of white clover plants

For the Tr-IFS2_12 transformation, 22 regenerants from 13 individual plants regenerated, were transferred into a greenhouse and grew into mature plants. Some regenerants grew two or three shoots which were grown further. Genomic DNA extracted from the plants was tested for the transgene insert where a band of the predicted size (2041 bp) indicated successful transformation. The negative control (-) was wild type white clover plant 6525/5 DNA and the positive control (+) was pRSH1+IFS2_12 vector DNA. Thirteen plants from eight individuals were PCR positive while eight plants from six individuals were PCR negative, indicating that they were wild type escapes from the selection process. Two plants grew from regenerant P07, one was PCR positive (P07B) and the other PCR negative (P07A) (Figure 4.6A). Sixteen of the Tr-IFS2_12 regenerated plants growing in the greenhouse were re-tested 21 months later. Of the 16 plants, 15 gave the same PCR result as previously (Figure 4.6B). Only plant P08B tested PCR positive in 2006 and PCR negative in 2007.

Twelve IFS/CHI regenerants, and regenerants from their wild type twin cotyledons, were grown and tested for the presence of the transgene. Six of these IFS/CHI plants tested PCR positive for the transgene with a band of 144 bp and all wild type plants tested negative (Figure 4.6C).

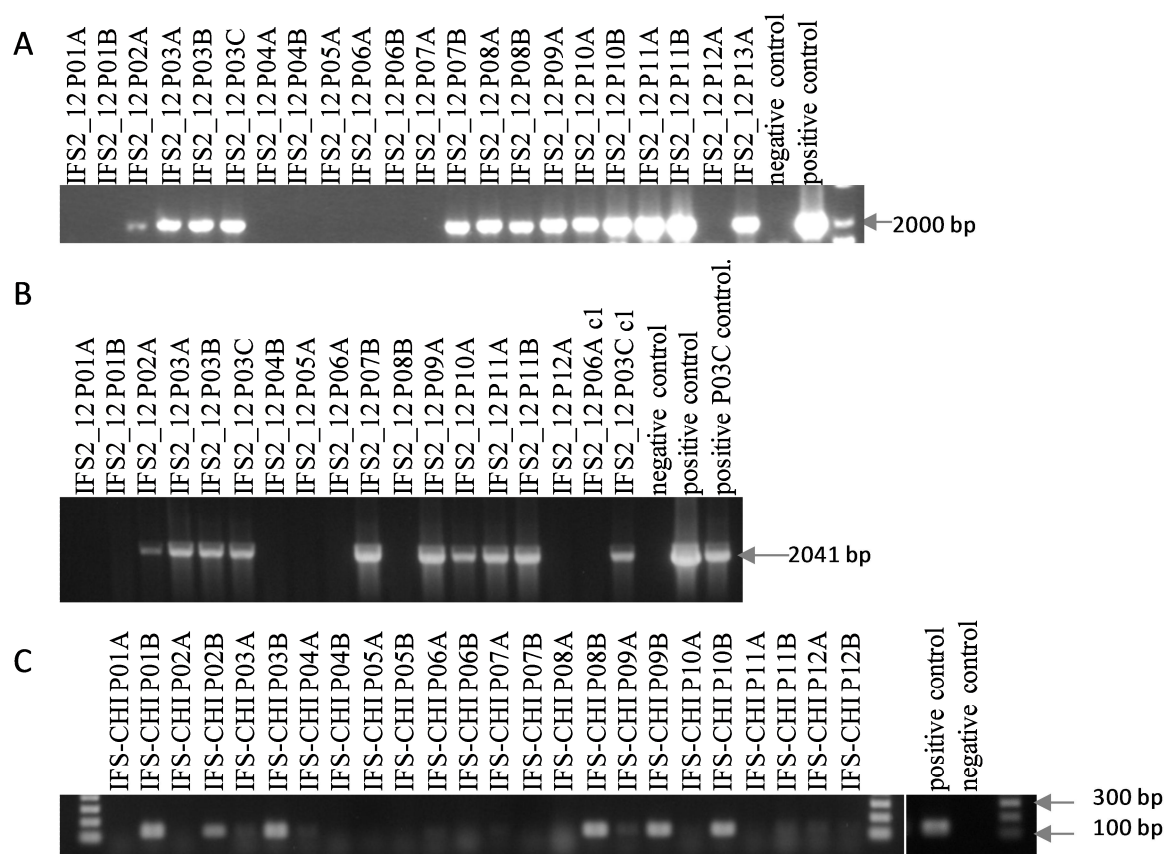


Figure 4.6 PCR tests for the transgene in white clover IFS2_12 (A, B) and IFS/CHI (C) transformants. A, initial test of IFS2_12 regenerants; B, retest of IFS2_12 regenerants 21 months later; C, test of IFS/CHI regenerants. In Tr-IFS2_12, the A to C letters represent different regenerants from the same explant while for Tr-IFS/CHI, A represents the wild-type clone while B represents the transgenic plant.

4.3.4 Isoflavonoid content and IFS gene expression in Tr-IFS2_12 transgenic white clover

The detection of isoflavonoids in HPLC chromatograms was only possible for formononetin glucoside malonate because the other compounds were below the detection limit of the instrument. Formononetin glucoside malonate was the isoflavonoid found in the highest concentration, and made up the majority of the total isoflavonoids, in white clover.

Formononetin glucoside malonate content in leaves and roots ranged from 528 to 4972 $\mu\text{g g}^{-1}$ DW and was consistently higher in the roots than in the shoots of all individuals sampled. The range of formononetin contents of the wild type plants was similar to that of the transformed plants and no clear pattern distinguishing the transgenic plants from the wild type plants is evident (Figure 4.7A).

The quantity of IFS mRNA transcripts in 25 ng of RNA was highly variable across the 22 samples, varying 5,000 fold between the highest (P03C) and the lowest (P06A) (Figure 4.7B). Within the two groups of wild-type and transformed plants there was also considerable variability, ranging from 109 to 58,700 transcripts in the wild-type plants and from 5,940 to 512,000 transcripts in the transformed plants. In samples P11 and P03, where two or three clones of one plant genotype were tested, the number of IFS transcripts of the clones was similar. The two clones of P01 had numbers of IFS transcripts that varied 3-fold. Between the two groups there was a trend for the transformed plants to have a higher number of IFS transcripts than non-transformed plants. Six of the nine transformed plants had more IFS transcripts than the highest non-transformed plant (P01B). The average number of IFS transcripts detected in the three P03 regenerants was eleven times higher than the average of all of the wild-type plants.

When compared with the *maximum* wild-type IFS expression of 58,700 transcripts per 25 ng of RNA (in P01B), the P10A, P02A and P03 averages were two, three and six times the level of P01B, respectively. When compared with the *average* wild-type IFS expression, P10A had 4.5 times the level of IFS transcripts, P02A had 6.3 times and for P03 the level was 12.9 times. That equates to 4.7, 7.1 and 16.0 wild-type standard deviations higher than the wild-type average for P10A, P02A and P03 average, respectively. In all cases the difference is at $P=0.01$.

The correlation between IFS transcript levels and formononetin content in white clover was $R^2=0.1875$ for Tr-IFS2_12 leaves and $R^2=0.1954$ for wild-type leaves.

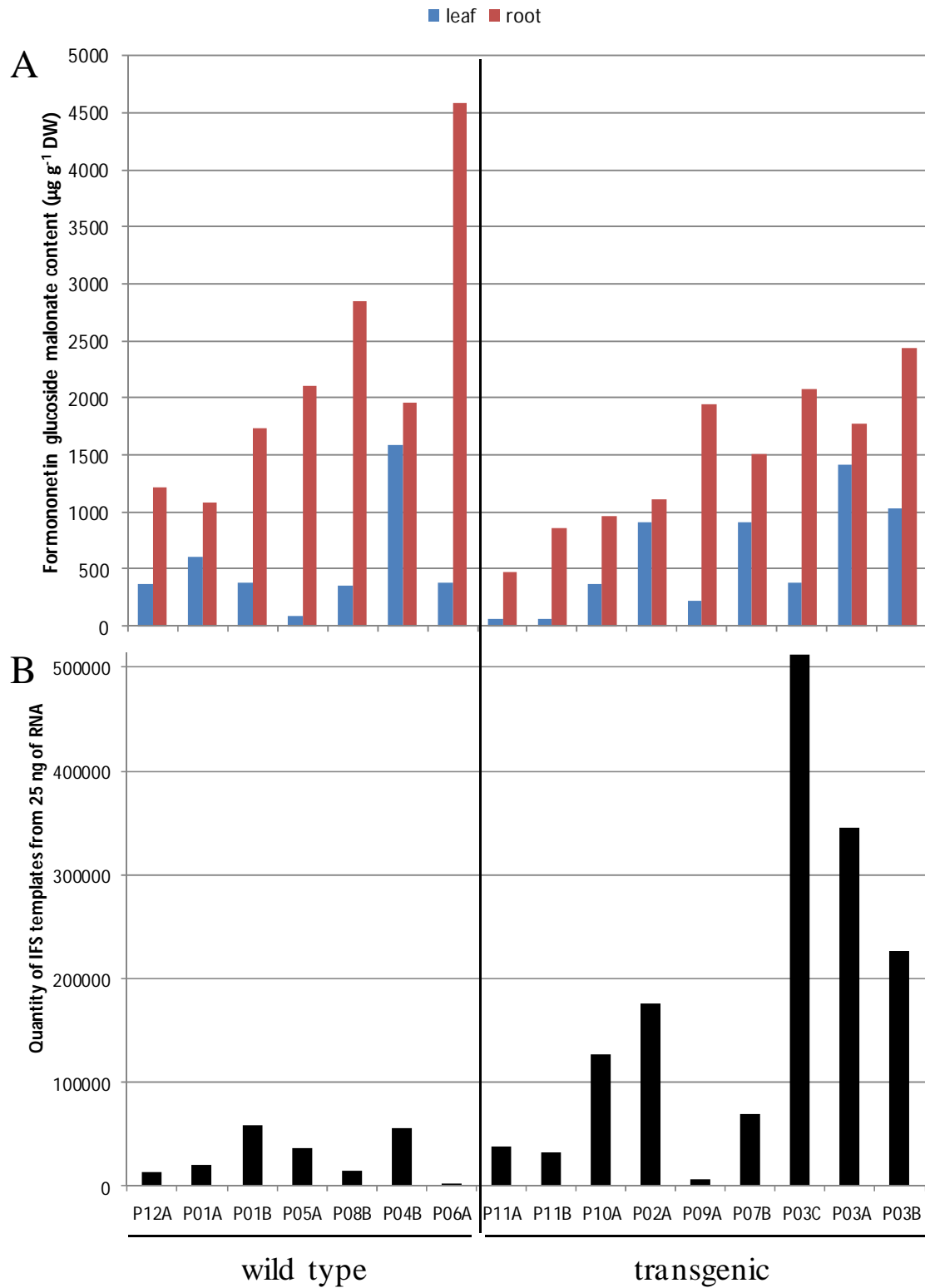


Figure 4.7 Formononetin content of the leaves and roots (A), and expression of IFS in the leaves (B), of Tr-IFS2_12 regenerants.

4.3.5 Isoflavonoid content of Tr-IFS/CHI transgenic white clover

The six Tr-IFS/CHI regenerants and their wild-type clones had formononetin glucoside malonate levels lower than those measured in the Tr-IFS2_12 regenerants, ranging from 70 to 307 $\mu\text{g g}^{-1}$ DW (Figure 4.8). The wild-type plants had average of 141 $\mu\text{g g}^{-1}$ DW formononetin with an and a standard deviation of 27 $\mu\text{g g}^{-1}$ DW. The transgenic plants fell into three groups. Plant P01's formononetin content was 46% of the wild-type clone. Plants P02, P08 and P10 had levels of formononetin within 12% of the wild type. Plants P03 and P09 had levels of 160 and 270% of their wild type clone, respectively.

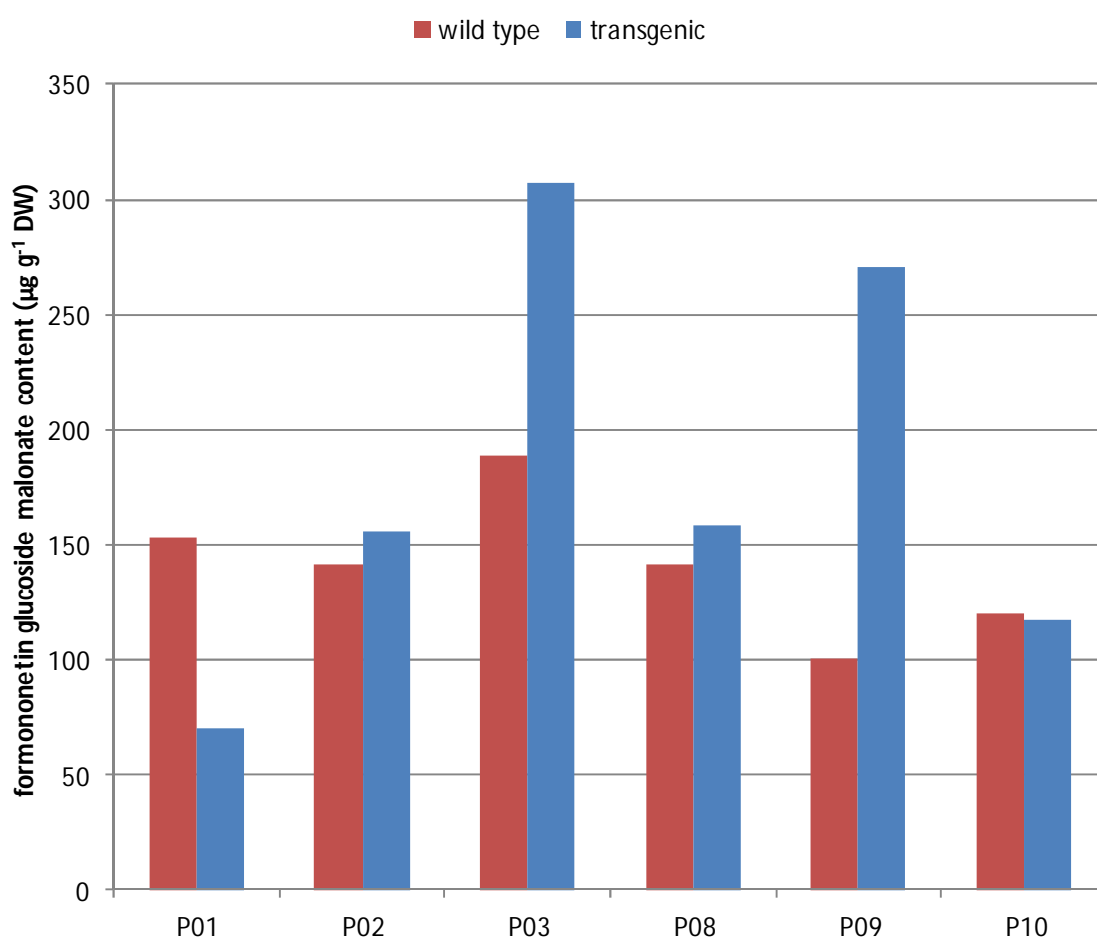


Figure 4.8 Formononetin glucoside malonate content of Tr-IFS/CHI transgenic plants and their wild-type clones.

4.4 Discussion

Analysis of the natural levels of isoflavonoids in white and red clover showed that red clover var Pawera contains 50 to 100 times more formononetin than white clover and that, while each cultivar has a characteristic concentration of individual isoflavonoids, the variability between individuals within the species can be quite high – the standard deviation of the white clover varieties ranged from 30% to 127% of the average. Two chemotypes were discovered in red clover ‘Pawera’ and it would be interesting to determine if these chemotypes were genetic or caused by environmental factors such as fungal infection. The patterns of isoflavonoid distribution in different tissues of four red clover cultivars shows the genetic variation available for this trait.

Biochanin A levels were quite consistent in both red clover cultivars (standard deviations of 24% for ‘Pawera’ and 10% for ‘G27’) while formononetin levels had been reduced tenfold in ‘G27’. This suggests that clover cultivars can be produced, where the level of one isoflavonoid is changed, while another remains the same.

The expression of IFS is positively correlated to the isoflavonoid levels within white clover in the initial experiments but the detected IFS expression levels of red clover were well below that of the white clover plants. This could be explained if the red clover IFS gene sequences are different to those of white clover and the primers that were used did not anneal well to the red clover cDNA. The red clover IFS genes from the two cultivars were not cloned so it is not known if the sequences vary from white clover and this would need to be tested in future experiments.

Transgenic white clover plants expressing IFS2-12 and IFS/CHI were produced in two consecutive transformation experiments. The Tr-IFS2_12 regenerants were PCR tested twice for the presence of the transgene, 21 months apart, and 15 out of 16 tested the same as previously, with only P08B changing from PCR positive to PCR negative. This gives confidence that the transgene is stable in the plant and that all of the plant tissue is transgenic.

The white clover transformation method using seed from the 'Huia' cultivar as explants was time consuming and had a low efficiency, resulting in a limited number of confirmed transgenic plants. In *A. tumefaciens*-based plant transformation, the transgenes integrate into the plant genome at random positions and possibly as repeats (Peach and Velten, 1991). The number of transgenes and the location of integration affect the expression of the transgenes which may be expressed at less than the maximal rate and can also be suppressed by transcriptional or post-transcriptional gene silencing (Stam et al., 1997). This means that each transformant must be analysed as an individual. The formononetin levels and IFS transcript numbers of the wild-type plants can be averaged to provide an average 'baseline level', for the 'Huia' cultivar, that the transgenic plants can be compared with. Since the white clover transformation system uses different genotypes as explants, and the outcrossing white clover cultivars are genetically diverse to reduce inbreeding depression, there can be large phenotypic variations between individual genotypes of the cultivar that are transformed. This limits the utility of the 'baseline level' measurements to outcomes where the transgene causes a large change in the phenotype of the transgenic plant. Moderate or small changes in the phenotype of the transgenic plant will not be statistically significantly different to the highly variable wild type levels with their large standard deviation.

Prior to the Tr-IFS2_12 transformation it was assumed that the transgene would cause a large increase in isoflavonoid levels, sufficiently large to be statistically significant. The results of the transformation had a large range of formononetin levels (Figure 4.7A). There was an equally wide range of formononetin levels in both the wild-type and transgenic plants. Thus no change in formononetin levels between the wild-type and the transgenic Tr-IFS2_12 plants can be confirmed.

While the IFS transcript number in the Tr-IFS2_12 leaves were very variable, there was a difference between the values of three transformants - P10A, P02A, and the average values of the three P03 regenerants, over the range of the wild-type levels. The conclusion drawn is that the IFS expression in these three transgenic genotypes was increased over the wild-type levels by the expression of the transgene.

The correlation between IFS transcript levels and formononetin content in white clover was very low for Tr-IFS2_12 leaves and also for wild-type leaves. This does not support the

hypothesis that increasing IFS expression will lead to increased levels of isoflavonoids and is evident by the isoflavonoid contents of regenerants P10A, P02A and P03A to C, which were within the range of the levels found in the wild-type controls.

Given that the natural genetic variability of white clover cultivars limits the identification of small or moderate changes in isoflavonoid levels, a second transformation experiment was carried out which controlled for this genetic variation by growing and testing wild-type clones of all transgenic plants to act as individual baseline controls for each transgenic plant. The novel twinned cotyledon transformation system took advantage of the fact that there were two genetically identical explants from each dissected seed and that white clover regenerates into new plants at high efficiency when not under AMM selection. The two cotyledons from each seed were tracked by their position in agar tissue culture media and the high regeneration rate of wild type seed under no herbicide selection meant that almost all putative transgenic regenerants had their wild-type clone regenerate into plants as well.

For the second transformation, IFS/CHI was chosen as a transgene because it had been shown to be functionally active in model plants (Tian and Dixon, 2006). At the time, the white clover IFS genes isolated in this project (Chapter 2) were still in the process of being tested for functionality in tobacco and the lack of isoflavonoid increase in the transgenic Tr-IFS2_12 plants did not support classifying it as functional either. It was thus uncertain, at the time, that IFS2_12 was functionally active, although this was subsequently confirmed as described in Chapter 3. Additionally, the CHI component of IFS/CHI was designed to increase the production of the precursors to genistein and daidzein and thus to increase isoflavonoid production in transgenic plants over that of plants only overexpressing IFS. This was the case in transgenic tobacco (Tian and Dixon, 2006), therefore this gene was chosen, to maximise the potential production of isoflavonoids.

The formononetin glucoside malonate levels of the six Tr-IFS/CHI plants were compared with those of their wild type clones and three groups of plants were identified – P01 had lower levels than the wild type, P02, P08 and P10 had very similar levels of formononetin compared to the wild type and P03 and P09 had higher levels than the wild-type (Figure 4.8). Assuming that the standard deviation of the individual clones is the same as the standard deviation measured for all of the wild type plants, P02, P08 and P10 were not

different from their wild type plants while the formononetin levels of P01, P03 and P09 were -3.0, 4.4 and 6.2 standard deviations different from their wild type controls. It is thus likely that the changes in isoflavonoid levels were caused by the transgene. It is possible that the increases in formononetin glucoside malonate levels in plants P03 and P09, compared with the inconclusive results of the Tr-IFS2_12 experiment, were due to the inclusion of CHI in the Tr-IFS/CHI plants. If CHI is functioning as intended, then the higher production of the precursors naringenin and liquiritigenin would lead to higher production of isoflavonoids, presumably mainly daidzein in this case, which would then be processed further to make formononetin glucoside malonate.

These findings have to be viewed with caution because of the limitations in the IFS transcript and isoflavonoid measurements in these experiments. While the qPCR primers for IFS were designed to bind to all cloned white clover sequence variants, there is a possibility, as discussed in Chapter 2, that further sequence variants are expressed that do not match the sequence of the primer and these would not be amplified and included in the IFS expression measurement. The IFS sequences of the wild-type red clover genotypes that were tested (results in Figure 4.5) may not match the primer designed for white clover and this would explain why the measured red clover IFS expression levels were lower than those of white clover. Additionally, the lack of a housekeeping gene to normalise the IFS expression against may be introducing additional variability into the IFS expression results. The samples were normalised according to total RNA concentration.

The isoflavonoid extracts were analysed by HPLC which had a detection limit greater than the quantities of most of the isoflavonoids from white clover extracts. This limited the analysis of isoflavonoids to formononetin glucoside malonate which is the major isoflavonoid found in white clover (Section 4.3.1). Given this, it is assumed that in white clover there is CHR, IOMT, glycosyltransferase and malonyltransferase activity that drives isoflavonoid production towards formononetin glucoside malonate. The level of formononetin and its conjugates is taken as being representative of the overall isoflavonoid level in white clover. The product of both IFS2_12 and IFS/CHI is genistein and daidzein which were below HPLC detection limits and could not be measured.

To address these concerns, a larger experiment was designed and carried out with six replicates of each Tr-IFS/CHI plant. The qPCR reactions of these included actin and PP2A housekeeping genes for data normalisation and tested different primers, some of which were designed to be specific to some IFS sequence variants to try to assess the expression levels of different sequence variants. The isoflavonoid extracts were analysed by LCMS to allow the measurement of much smaller quantities of isoflavonoids and the discrimination between all of the major types and conjugates of isoflavonoids. Unfortunately technical problems (inhibition of reverse transcription is suspected) resulted in qPCR IFS expression results that were abnormally low and the LCMS measurement was set up for aglycone measurement so the isoflavonoid conjugates were not measured. Due to time constraints, this experiment was not repeated.

It is not possible to confidently state why the levels of IFS transcripts and formononetin glucoside malonate varied so much in both the wild-type and the Tr-IFS2-12 transgenic plants without further investigation into the natural regulation of isoflavonoid biosynthesis and into the position, copy number and possible gene silencing of the transgenes in the white clover genome of the individual transgenic plants. However, the results show that Tr-IFS2_12 plants P10A, P02A and P03A to C had higher levels of IFS transcripts than the wild type controls without causing an increase in isoflavonoid levels. This does not support the hypothesis that higher IFS expression will increase isoflavonoid levels. On the other hand, the Tr-IFS/CHI plants P03 and P09 had higher levels of formononetin than their clonal controls which suggests that this transgene had a positive effect on isoflavonoid levels in these two transformants.

5 General discussion

The purpose of this project was to investigate metabolic engineering of isoflavonoids in white clover. The optimal outcome of this would be the ability to change the quantities and types of isoflavonoids so that pest and disease resistance and the production of root nodules were increased while avoiding the production of high levels of estrogenic isoflavonoids that cause a reduction in livestock fertility. The hypotheses tested were that IFS genes cloned from white clover are both necessary and sufficient for producing genistein in tobacco plants that are expressing these as transgenes, and that increased expression of IFS will increase the concentration of isoflavonoids in transgenic white clover.

To clone isoflavonoid biosynthesis genes from white clover, the Gene-Thresher (Warek et al., 2005) white clover genomic DNA database was searched for sequences similar to those available in public databases. This resource allowed the rapid creation of contigs of white clover sequences that primers could be designed from. However, the depth of coverage of the database was not complete so the IOMT contig was not full-length and the IFS contig was a chimera of two sequence variants (Figure 2.1). However, a full length contig of one sequence variant was assembled for CHR. The Gene-Thresher contigs were more closely related to the white clover genes cloned in this project than most publically available sequences, increasing the chances of designing a functioning primer for PCR amplification of the white clover genes. Public sequences were still required for designing IOMT primers and these successfully amplified full length IOMT genes from white clover so the same approach could have been taken for the other genes. The PCR-amplification of genes from white clover appeared to limit the number of sequence variants that were cloned as seen in Figure 2.14, which shows that the white clover genes cloned in this study were genetically distinct to those cloned by Jung et al. (2000). Once a complete genome sequence of white clover is published, this will be a more useful resource for designing primers to amplify any gene. This is particularly true if the genomic DNA of the plant that was sequenced was used as the PCR template – something that was not possible with the Gene-Thresher database since the DNA of the sequenced genotype was not available.

Functional testing of cloned genes by expressing them in transgenic tobacco had been carried out previously (Yu et al., 2000; Tian and Dixon, 2006) and the results of this study are similar – expression of IFS or IFS/CHI produced low levels of genistein and genistein conjugates (Figure 3.14, Figure 3.16), and the levels of these increased in tissues that produced anthocyanins (Figure 3.15) (Grotewold et al., 1998). This confirms the functionality of the IFS/CHI and IFS2_12 genes and confirms the hypothesis that these genes are necessary and sufficient for producing genistein in a naïve model plant.

The use of anthocyanin-producing, ANT1-expressing tobacco plants as test plants for leaf infiltration was novel and the generally increased genistein levels showed that a greater presence of precursors produced greater amounts of genistein. This was done to maximise the quantity of isoflavonoids, since the levels being produced were below the detection limit of HPLC. Isoflavonoid levels were greater in coloured plants but the effect seemed to be only moderate, the progeny of Tr-IFS2_12 plant 23722 produced $6.6 \mu\text{g g}^{-1}$ DW of genistin without the presence of anthocyanins (Figure 3.17) and the use of the much more sensitive LCMS instrumentation obviated the need for increased levels for detection. Genistein was produced irrespective of whether ANT1 was stably transformed and IFS was infiltrated or IFS was stably transformed and ANT1 was infiltrated. The stable Nt-IFS/CHI transformants might have been expected to accumulate greater levels of isoflavonoids than when that gene was infiltrated, since the transgenic plants had a substantially longer time to accumulate isoflavonoids than the seven days that the infiltrated tissues are grown for, but this was not evident (Figure 3.15). No isoflavonoids other than genistein and its conjugates were detected in IFS-expressing transgenic tobacco tissues and no traces of genistein were detected in tobacco that did not contain an IFS transgene. This indicates that isoflavonoid biosynthesis genes such as IFS or IOMT, if present in the wild-type tobacco genome, were not being expressed. The use of leaf infiltration to test the functionality of isoflavonoid biosynthesis genes is novel and the quantities of genistein produced were similar to those in stable transformations, making this a viable functional testing technique that is faster than producing stable transformants.

The transformation of white clover to overexpress IFS or IFS/CHI was similar to the work of Deavours and Dixon (2005) who transformed alfalfa with constitutively expressed *M. truncatula* MtIFS1, and Jung et al. (2003) who expressed soybean IFS in soybean plants

and measured the isoflavonoid content of the seed. In both cases, as with white clover, the plants already contained and expressed isoflavonoid biosynthesis genes and produced isoflavonoids but in alfalfa isoflavonoids are not naturally produced in the seeds and in soybean the aim was to increase the isoflavonoid content in the seed. In both cases, overexpression of IFS produced moderately positive results – alfalfa leaves did produce isoflavonoids but at low levels and the only significant increase of isoflavonoids in soybean seeds was 1.7 times the wild-type level, in one transformant out of fourteen. These results are similar to the stable white clover transformations in this study – the IFS2_12 transformants increased IFS expression, as had happened in the transgenic alfalfa plants, but no increase of formononetin was evident (Figure 4.7). The IFS/CHI transformants did produce greater levels of formononetin in two of the six plants when compared with their wild-type clones (Figure 4.8).

The use of the wild-type clones in the twinned cotyledon transformation system was a novel method that allowed the discrimination of moderate changes in metabolite levels in a genetically diverse group of white clover transformants because it removed the genetic component of phenotypic variability. This leaves just the environmental component of phenotypic variability which should be minimised by growing all plants under the same conditions. Although the twinned cotyledon system is more time consuming, the greater power of discrimination between phenotypes is a valuable improvement of the current white clover transformation protocol. This method was used for the transformation of white clover with the IFS/CHI construct.

As with model plants (Yu et al., 2000), higher production of isoflavonoids in legumes generally requires an increase in precursor accumulation through the increased production of the precursor or blockage of a competing biochemical pathway (Yu et al., 2003). CHI produces the precursors for IFS and this can be a rate-limiting step (Tian and Dixon, 2006). The IFS/CHI gene was designed to increase the availability of isoflavonoid precursors to the IFS enzyme and thus increase isoflavonoid levels greater than those caused by the expression of IFS alone. The only increase in formononetin in transgenic white clover was when expressing the IFS/CHI gene. However, these results cannot be compared to the results of the IFS2_12 transformation experiment since the latter cannot account for the genetic

variability of the transformed individuals. It thus remains uncertain whether IFS/CHI is more effective than IFS2_12 at increasing isoflavonoid levels in white clover.

The increase of formononetin in the two Tr-IFS/CHI transformants could be explained, at least in part, by the increase in IFS expression caused by the expression of the transgene. Hence these two plants would represent the desired outcome of the transformation. To test this, measurement of IFS expression and isoflavonoid content of replicates of these plants should be carried out. This work was done but technical problems caused the failure of the expression analysis and time constraints prevented a repetition of the experiment. Currently, the hypothesis that increased expression of the IFS gene (alone) will increase the concentration of isoflavonoids in transgenic white clover is not supported by these results from transgenic white clover plants and the relation between IFS expression and isoflavonoid content in white clover remains unknown.

Future directions

Two potentially useful methods of producing more optimal isoflavonoid levels in the future are marker-assisted selection (Collard and Mackill, 2008) and transformation of plants with isoflavonoid transcription factors. However, the transcription factors that control isoflavonoid biosynthesis have not yet been discovered. If found, these would be very promising candidates for transgene expression in legumes to manipulate isoflavonoid content. Transcription factors can activate several genes at once and have evolved to efficiently activate whole biosynthesis pathways in a coordinated manner (Broun, 2004). The use of transcription factors is likely to be more effective at increasing isoflavonoid levels than overexpressing single, or even multiple, genes coding for individual isoflavonoid biosynthesis enzymes. To increase soybean isoflavonoid content further, Yu et al. (2003) expressed the maize C1 and R transcription factors to increase precursors for the endogenous IFS to produce isoflavonoids and found an approximately 2-fold increase in overall isoflavonoid levels and a greatly altered ratio of genistein: daidzein, possibly because the transcription factors substantially increased CHR transcription. As the transcription factors for white clover anthocyanin production have not yet been isolated, a heterologous system similar to that used by Yu et al. (2003) could be attempted, to increase isoflavonoid levels. When flavanone 3-hydroxylase expression was suppressed in addition to the maize C1 and R

transcription factors, to block the anthocyanin pathway, isoflavonoid levels in soybean seed increased up to 4-fold greater than wild-type. This method of increasing isoflavonoids may be unsuitable in white clover tissues because it blocks the production of anthocyanins and other flavonoids and this may have adverse effects on the plants deprived of these secondary metabolites.

Genetic mapping of isoflavonoid genes could be carried out using putative SNPs discovered in Chapter 2 as long as some of the SNPS are present in one allele (a gene found in one location in the genome) and not just between sequence variants from different areas of the genome. Such SNPs could be discovered by cloning and sequencing a number of genes from two genetically diverse mapping population parents and designing allele-specific assays that can be tested in a mapping population (e.g. Barrett et al., 2004). An alternative would be to quantify the individual and total isoflavonoids in a white clover mapping population that segregates for these phenotypes and to map these variations as quantitative trait loci (Doerge, 2002). While genetic markers would be useful for marker-assisted breeding, quantitative trait loci for isoflavonoid content may map to regulatory elements of isoflavonoid biosynthesis and this would narrow down the search for transcription factors to a small proportion of the genome. Homology-based searches for transcription factors in the forthcoming genome sequence could be narrowed down, using this mapping information, to a manageable number of candidates that can then be functionally tested in white clover or a more easily transformed model legume.

An important future direction would be to study the interaction between isoflavonoids, root nodulation and nitrogen fixation. Given that nitrogen fixation is arguably the most important role for white clover in temperate pastures, it is an area of research that could yield great benefits in increasing nitrogen production in clover, perhaps by breaking the negative feedback that inorganic soil nitrogen levels have on isoflavonoid and root nodule production in the roots (Subramanian et al., 2006), so that increased amounts of nitrogen can be fixed for the benefit of increasing the productivity of the whole pasture ecosystem. Since nitrogen fixation is metabolically costly, the current negative feedback loop is useful for individual white clover plants which have limited resources to grow and are competing with grass in existing pastures. However, it may be desirable to increase nitrogen fixation in white clover, even at the expense of the yield of white clover fodder. A more persistent white clover plant

that fixes more nitrogen may produce higher overall pasture dry matter yield through the greater growth of the grass component of the pasture and thus increase animal productivity.

Given the diverse biological activity of isoflavonoids in plants and animals, one should aim to balance the advantages of insect resistance and root nodulation with the potential disadvantages of estrogenicity in animal. In a white clover plant that was metabolically engineered to produce optimal isoflavonoid levels, the concentration, location and types of isoflavonoids would need to be carefully considered. An ideal isoflavonoid profile in white clover, given our current state of knowledge, would be one that increases genistein and daidzein content in, and exudation from, the roots to encourage a greater production of root nodules. The levels of biochanin A would be high in the leaves (about 1.5% to 4% of DW) to reduce feeding by CRW and other leaf-eating insects. Formononetin would be present at moderate levels (less than 0.5% of DW) in the leaves to provide some protection from insect feeding while being low enough to have little or no effect on the fertility of livestock. For some grazing animals that are less sensitive to formononetin than ewes, the leaf formononetin levels could be increased to improve insect resistance further, balanced with a concomitant decrease in biochanin A so that isoflavonoid production does not place an excessive drain on the plant's resources. Root formononetin content would be relatively high and the production of more formononetin would be quickly inducible when damage by nematodes and root-feeding larvae is detected. More potent isoflavonoid phytoalexins such as medicarpin would be produced or increased to provide resistance to fungal infection.

The science of metabolic engineering of isoflavonoids in plants is still developing, and future knowledge and technology is likely to enable the metabolic engineering of more optimal isoflavonoid levels in plants to reduce pests and diseases and increase nitrogen fixation without causing reduction of fertility in livestock. This research project has laid the groundwork for such metabolic engineering in white clover that can be built upon in the future.

References

- Adams NR** (1995) Detection of the effects of phytoestrogens on sheep and cattle. *J Anim Sci* **73**: 1509-1515
- Akashi T, Aoki T, Ayabe S** (1998) Identification of a cytochrome P450 cDNA encoding (2S)-flavanone 2-hydroxylase of licorice (*Glycyrrhiza echinata* L.; Fabaceae) which represents licodione synthase and flavone synthase II. *FEBS Lett* **431**: 287-290
- Akashi T, Aoki T, Ayabe S** (1999) Cloning and Functional Expression of a Cytochrome P450 cDNA Encoding 2-Hydroxyisoflavanone Synthase Involved in Biosynthesis of the Isoflavonoid Skeleton in Licorice. *Plant Phys* **121**: 821–828
- Akashi T, Aoki T, Ayabe S** (2005) Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis. *Plant Physiol* **137**: 882–891
- Akashi T, Sawada Y, Aoki T, Ayabe S** (2000) New scheme of the biosynthesis of formononetin involving 2,7,4'-trihydroxyisoflavanone but not daidzein as the methyl acceptor. *Biosci Biotechnol Biochem* **64**: 2276–2279
- Akashi T, Sawada Y, Shimada N** (2003) cDNA cloning and biochemical characterization of S-adenosyl-L-methionine: 2,7,4'-trihydroxyisoflavanone 4'-O-methyltransferase, a critical enzyme of the legume isoflavonoid phytoalexin pathway. *Plant Cell Physiol* **44**: 103–112
- Albert NW, Lewis DH, Zhang H, Irving LJ, Jameson PE, Davies KM** (2009) Light-induced vegetative anthocyanin pigmentation in *Petunia*. *J Exp Bot* **60**: 2191–2202
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402
- Ayabe S, Udagawa A, Furuya T** (1988) Stimulation of chalcone synthase activity by yeast extract in cultured *Glycyrrhiza echinata* cells and 5-deoxyflavanone formation by isolated protoplasts. *Plant Cell Rep* **7**: 35-38

Ballance GM, Dixon RA (1995) *Medicago sativa* cDNAs Encoding Chalcone Reductase. *Plant Physiol* **107**: 1027-1028

Barrett B, Griffiths A, Schreiber M, Ellisson N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, Bryan G, Woodfield D (2004) A microsatellite map of white clover. *Theor Appl Genet* **109**: 596-608

Beck, AB (1964) The oestrogenic isoflavones of subterranean clover. *Aust J Agric Res* **15**: 223-230

Bednar RA, Hadcock JR (1988) Purification and characterization of chalcone isomerase from soybeans. *J Biol Chem* **263**: 9582–9588

Beggs CJ, Stolzer-Jehle A, Wellmann E (1985) Isoflavonoid formation as an indicator of UV stress in bean (*Phaseolus vulgaris* L.) leaves. *Plant Physiol* **79**: 630–634

Bendahmane A, Querci M, Kanyuka K, Baulcombe DC (2000) *A. tumefaciens* transient expression system as a tool for isolation of disease resistance genes: application to the Rx2 locus in potato. *Plant J.* **21**: 73–81

Bennetts HW, Underwood EJ, Shier FL (1946) A specific breeding problem of sheep on subterranean clover pastures in Western Australia. *Aust Vet J* **22**: 2-12

Bernard P, Couturier M (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol* **226**: 735-745

Bhattacharyya MK, Ward EWB (1987) Biosynthesis and metabolism of glyceollin I in soybean hypocotyls following wounding or inoculation with *Phytophthora megasperma* f. sp. Glycinea. *Physiol Mol Plant Pathol* **31**: 387–405

Bisby FA, Buckingham J, Harborne JB (Eds) (1994) *Phytochemical dictionary of the Leguminosae, Vol I. Plants and their constituents.* Chapman and Hall, New York

Blount JW, Dixon RA, Paiva NL (1993) Stress responses in alfalfa (*Medicago sativa* L.) XVI. Antifungal activity of medicarpin and its biosynthetic precursors; implications for the genetic manipulation of stress metabolites. *Physiol Mol Plant Pathol* **41**: 333–349

- Blyden ER, Doerner P, Lamb CJ, Dixon RA** (1991) Sequence analysis of a chalcone isomerase cDNA of *Phaseolus vulgaris*. *Plant Mol Biol* **16**: 167-169
- Bomati EK, Austin MB, Bowman ME, Dixon RA, Noel JP** (2005) Structural elucidation of chalcone reductase and implications for deoxychalcone biosynthesis. *J Biol Chem* **280**: 30496–30503
- Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C** (2000) Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. *Plant Cell* **12**: 2383-2394,
- Braun E, Grotewold E** (1999) Newly Discovered Plant c-myb-Like Genes Rewrite the Evolution of the Plant myb Gene Family. *Plant Phys* **121**: 21–24
- Bridge MA, Klarman WL** (1973) Soybean phytoalexin, hydroxyphaseollin, induced by ultraviolet irradiation. *Phytopathol* **63**: 606–609
- Broughton WJ, Jabbouri S, Perret X** (2000) Keys to symbiotic harmony. *J Bacteriol* **182**: 5641–5652
- Broun P** (2004) Transcription factors as tools for metabolic engineering in plants. *Curr Opin Plant Biol* **7**: 202-209
- Bruce W, Folkerts O, Garnaat C, Crasta O, Roth B, Bowen B** (2000) Expression profiling of the maize flavonoid pathway genes controlled by estradiol-inducible transcription factors CRC and P. *Plant Cell* **12**: 65–79
- Byers RA, Kendall WA** (1982) Effects of plant genotypes on growth and survival of *Sitona* spp. larvae. *Environ Entom* **11**: 440-443
- Caldwell MM, Robberecht R, Flint S** (1983) Internal filters: prospects for UV-acclimation in higher plants. *Physiol Plant* **58**: 445–450
- Caradus JR, Woodfield DR, Stewart AV** (1996) Overview and vision for White Clover. In Woodfield DR, ed, *White Clover: New Zealand's Competitive Edge*. Proceedings of a joint symposium between the Agronomy Society of NZ and NZ Grassland Association, Lincoln University NZ, 21-22 November 1995

Collard BCY, Mackill DJ (2008) Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci* **363**: 557-572

Cook R, Tiller SA, Mizen KA, Edwards R (1995) Isoflavonoid metabolism in resistant and susceptible cultivars of white clover infected with the stem nematode *Ditylenchus dipsaci*. *J Plant Physiol* **146**: 348-354

Cooper BM, Crush JR, Woodfield DR, Willoughby BE, Panckhurst KA (2003) Red clover – a legume showing tolerance to clover root weevil. *Proc N Z Grassland Assoc* **65**: 153-156

Cosgrove GP, Burke JL, Death AF, Lane GA, Fraser K, Pacheco D (2006) The effect of clover-rich diets on cows in mid lactation: production, behaviour and nutrient use. *Proc NZ Grassland Assoc* **68**: 267-273

Crush JR, Ouyang L, Gerard, PJ, Rasmussen R (2007) Effect of clover root weevil larvae on four annual forage legumes. *Proc NZ Grassland Assoc* **69**: 213-217

Dakora FD, Joseph CM, Phillips DA (1993) Alfalfa (*Medicago sativa* L.) root exudates contain isoflavonoids in the presence of *Rhizobium meliloti*. *Plant Physiol* **101**: 819-824

Dakora FD, Philips DA (1996) Diverse functions of isoflavonoids in legumes transcend anti-microbial definitions of phytoalexins. *Physiol Mol Plant Pathol* **49**: 1-20

de Vetten N, Quattrocchio F, Mol J, Koes R (1997) The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev* **11**: 1422–1434

Deavours B, Dixon R (2005) Metabolic Engineering of Isoflavonoid Biosynthesis in Alfalfa. *Plant Physiol* **138**: 2245-2259

Deavours BE, Liu C-J, Naoumkina MA, Tang Y, Farag MA, Sumner LW, Noel JP, Dixon RA (2006) Functional analysis of members of the isoflavone and isoflavanone O-methyltransferase enzyme families from the model legume *Medicago truncatula*. *Plant Mol Biol* **62**: 715–733

- Dewick PM** (1994) Isoflavonoids. In Harbourne JB (Ed) The Flavonoids, advances in research since 1986. Chapman and Hall, London. pp 117-238
- Dhaubhadel S, Farhangkhoei M, Chapman R** (2008) Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. J Exp Bot **59**: 981–994
- Dixon RA** (2004) Phytoestrogens. Ann Rev Plant Biol **55**: 225–61
- Doerge RW** (2002) Multifactorial genetics: Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet **3**: 43-52
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A** (2010) Geneious v5.3. Available from <http://www.geneious.com>
- Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP** (1999) Structure of chalcone synthase and the molecular basis of plant polyketide 7+ biosynthesis. Nat Struct Biol **6**: 775–784
- Geigert J, Stermitz FR, Johnson G, Maag DD, Johnson DK** (1973) Two phytoalexins from sugarbeet (*Beta vulgaris*) leaves. Tetrahedron **29**: 2703-2706
- Gerard PJ, Crush JR, Hackell DL** (2005) Interaction between *Sitona lepidus* and red clover lines selected for formononetin content. Ann Appl Biol **147**: 173-181
- Graham TL, Graham MY** (1991) Glyceollin elicitors induce major but distinct shifts in isoflavonoid metabolism in proximal and distal soybean cell populations. Mol Plant–Microbe Interact **4**: 60–68
- Grotewold E, Chamberlin M, Snook M, Siame B, Butler L, Swenson J, Maddock S, Clair GS, Bowen B** (1998) Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. Plant Cell **10**: 721–740
- Grotewold E** (2006) The Genetics and Biochemistry of Floral Pigments. Ann Rev Plant Biol **57**: 761–80

- Hadwiger LA, Schwochau ME** (1971) Ultraviolet light-induced formation of pisatin and phenylalanine ammonia lyase. *Plant Physiol* **47**: 588–590
- Hagmann M, Grisebach H** (1984) Enzymatic rearrangement of flavanone to isoflavone. *FEBS Lett* **175**: 199-202
- Hakamatsuka T, Mori K, Ishida S, Ebizuka Y, Sankawa U** (1998) Purification of 2-hydroxyisoflavanone dehydratase from the cell cultures of *Pueraria lobata*. *Phytochem.* **38**: 386-494
- Hancock K, Greig M, Bryan G** (2002) Milestone 21a Report. Pastoral Genomics Consortium, Auckland, New Zealand
- Hashim MF, Hakamatsuka T, Ebizuka Y, Sankawa U** (1990) Reaction mechanism of oxidative rearrangement of flavanone in isoflavone biosynthesis. *FEBS Lett* **271**: 219-222
- He X-Z, Dixon RA** (1998) Stress responses in alfalfa (*Medicago sativa* L). XXII. cDNA cloning and characterization of an elicitor-inducible isoflavone 7-O-methyltransferase. *Plant Mol Bio* **36**: 43–54
- He X-Z, Dixon RA** (2000) Genetic manipulation of isoflavone 7- O-methyltransferase enhances the biosynthesis of 4'-O-methylated isoflavonoid phytoalexins and disease resistance in alfalfa. *Plant Cell* **12**: 1689–1702
- Hernandez J, Heine G, Irani NG, Feller A, Kim M-G, et al.** (2004) Mechanisms of cooperation between MYB and HLH transcription factors in the regulation of anthocyanin pigmentation. *J Biol Chem* **279**: 48205–13
- Hofmann RW, Campbell BD, Bloor SJ, Swinny EE, Markham KR, Ryan KG, Fountain DW** (2003) Responses to UV-B radiation in *Trifolium repens* L. - Physiological links to plant productivity and water availability. *Plant Cell Environ* **26**: 603-612
- Horsch RB, Fry JE, Hoffmann NL, Eichholtz D, Rogers SG, Fraley RT** (1985) A Simple and General Method for Transferring Genes into Plants. *Science* **227**: 1229-1231

- Hsieh MC, Graham TL** (2001) Partial purification and characterization of a soybean β -glucosidase with high specific activity towards isoflavone conjugates. *Phytochem* **58**: 995-1005
- Jefferson RA, Burgess SM, Hirsh D** (1986) beta-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc Natl Acad Sci USA* **83**: 8447–8451
- Jez JM, Bowman ME, Dixon RA, Noel JP** (2000) Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nature Struct Biol* **7**: 786–791
- Jez JM, Bowman ME, Noel JP** (2002). Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. *Biochemistry* **41**: 5168–5176
- Jez JM, Noel J P** (2002). Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. *J Biol Chem* **277**: 1361–1369
- Johnson SN, Gregory PJ, Greenham JR, Zhang X, Murray PJ** (2005) Attractive properties of an isoflavonoid found in white clover root nodules on the clover root weevil. *J Chem Ecol* **31**: 2223-2229
- Jones P, Vogt T** (2001) Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta* **213**: 164–174
- Joshi CP, Chiang VL** (1998) Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant Mol Biol* **37**: 663–674
- Jung W, Yu O, Lau SC, O’Keefe DP, Odell J, Fader G, McGonigle B** (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nat Biotechnol* **18**: 208–212
- Jung WS, Chung IM, Heo H-Y** (2003) Manipulating Isoflavone Levels in Plants. *J Plant Biotechnol* **5**: 149-155
- Kape R, Parniske M, Brandt S, Werner D** (1992) Isoliquiritigenin, a strong nod gene- and glyceollin resistance-inducing flavonoid from soybean root exudate. *Appl Environ Microbiol* **58**: 1705-1710

Kapila J, de Rycke R, van Montagu M, Angenon G (1997) An *Agrobacterium*-mediated transient gene expression system for intact leaves. *Plant Sci* **122**: 101–108

Kaplan DT, Keen NT, Thomason IJ (1980a) Association of glyceollin with the incompatible response of soybean roots to *Meloidogyne incognita*. *Physiol Plant Pathol* **16**: 309–318

Kaplan DT, Keen NT, Thomason IJ (1980b) Studies on the mode of action of glyceollin in soybean incompatibility to the root knot nematode, *Meloidogyne incognita*. *Physiol Plant Path* **16**: 319–325

Kelly RW, Shackell GH, Allison AJ (1980) Reproductive performance of ewes grazing red clover (Grasslands Pawera) or white clover – grass pasture at mating. *NZ J Exp Agr* **8**: 87-91

Kimura Y, Aoki T, Ayabe S (2001) Chalcone Isomerase Isozymes with Different Substrate Specificities towards 6'-Hydroxy- and 6'-Deoxychalcones in Cultured Cells of *Glycyrrhiza echinata*, a Leguminous Plant Producing 5-Deoxyflavonoids. *Plant Cell Physiol* **42**: 1169-1173

Kochs G, Grisebach H (1986) Enzymic synthesis of isoflavones. *Eur J Biochem* **155**: 311–318

Krenn L, Unterrieder I, Ruprechter R (2002) Quantification of isoflavones in red clover by high-performance liquid chromatography. *J Chromatogr B* **777**: 123–128

Kuhn F, Oehme M, Romero F, Abou-Mansour E, Tabacchi R (2003) Differentiation of isomeric flavone/isoflavone aglycones by MS² ion trap mass spectrometry and a double neutral loss of CO. *J Mass Spectrom* **17**: 1941-1949

Landy A (1989) Dynamic, structural and regulatory aspects of lambda site-specific recombination. *Ann Rev Biochem* **58**: 913-949

Lane GA, Sutherland ORW, Skipp RA (1987) Isoflavonoids as insect feeding deterrents and antifungal components from roots of *Lupinus angustifolius*. *J Chem Ecol* **13**: 774-783

- Larkin JC, Walker JD, Bolognesi-Winfield AC, Gray JC, Walker AR** (1999) Allele-specific interactions between *ttg* and *gl1* during trichome development in *Arabidopsis thaliana*. *Genetics* **151**: 1591–1604
- Liu C-J, Blount JW, Steele CL, Dixon RA** (2002) Bottlenecks for metabolic engineering of isoflavone glycoconjugates in *Arabidopsis*. *Proc Natl Acad Sci USA* **99**: 14578–14583
- Liu CJ, Dixon RA** (2001) Elicitor-induced association of isoflavone O-methyltransferase with endomembranes prevents formation and 7-O-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. *Plant Cell* **13**: 2643–2658
- Lloyd AM, Walbot V, Davis RW** (1992) *Arabidopsis* and *Nicotiana* Anthocyanin Production Activated by Maize Regulators *R* and *C1*. *Science* **258**: 1773-1775
- Lockton S, Gaut BS** (2005) Plant conserved non-coding sequences and paralogue evolution. *TRENDS Genet.* **21**: 60-65
- Mathews H, Clendennen SK, Caldwell CG, Liu XL, Connors K, Matheis N, Schuster DK, Menasco DJ, Wagoner W, Lightner J, Wagner DR** (2003) Activation Tagging in Tomato Identifies a Transcriptional Regulator of Anthocyanin Biosynthesis, Modification, and Transport. *The Plant Cell* **15**: 1689–1703
- Meagher TR, Vassiliadis C** (2005) Phenotypic impacts of repetitive DNA in flowering plants. *New Phytologist* **168**: 71–80
- Miki T, Park JA, Nagao K, Murayama N, Horiuchi T** (1992) Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*: Mutants of DNA gyrase subunit A suppress *letD* (*ccdB*) product growth inhibition. *J Mol Biol* **225**: 39-52
- Mol JNM, Robbins MP, Dixon RA, Veltkamp E** (1985) Spontaneous and enzymic rearrangement of naringenin chalcone to flavanone. *Phytochem* **24**: 2267-2269
- Murashige T, Skoog F.** (1962) A revised medium for rapid growth and bio-assays with tobacco tissue cultures. *Physiol Plant* **15**: 473-497

- Murray PJ, Cheung AKM, Abberton MT** (2007) Intraspecific variation in *Trifolium pratense*: impact on feeding on host location by *Sitona lepidus* (Coleoptera, Curculionidae). *J Pest Sci* **80**: 51-57
- Nagashima S, Inagaki R, Kubo A, Hirotani M, Yoshikawa T** (2004) cDNA cloning and expression of isoflavonoid-specific glucosyltransferase from *Glycyrrhiza echinata* cell-suspension cultures. *Planta* **218**: 456–459
- Nair MG, Safir GR, Siqueira JO** (1991) Isolation and identification of vesicular-arbuscular mycorrhiza-stimulatory compounds from clover (*Trifolium repens*) roots. *Appl Environ Microbiol* **57**: 434-439
- Noel JP, Dixon RA, Pichersky E, Zubieta C, Ferrer J-L** (2003) Structural, functional, and evolutionary basis for methylation of plant small molecules. *Recent Adv Phytochem* **37**: 37–58
- Noguchi A, Saito A, Homma Y, Nakao M, Sasaki N, Nishino T, Takahashi S, Nakayama T** (2007) A UDP-glucose:isoflavone 7-O-glucosyltransferase from the roots of soybean (*Glycine max*) seedlings—purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis. *J Biol Chem* **282**: 23581-23590
- Paiva NL, Oommen A, Harrison MJ, Dixon RA** (1994) Regulation of isoflavonoid metabolism in alfalfa. *Plant Cell Tissue Organ Cult* **38**: 213-220
- Peach C, Velten J** (1991) Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. *Plant Mol. Biol.* **17**:49-60
- Piślewska M, Bednarek P, Stobiecki M, Zieliński M, Wojtaszek P** (2002) Cell wall-associated isoflavonoids and β -glucosidase activity in *Lupinus albus* plants responding to environmental stimuli. *Plant Cell Environ* **25**: 29-40
- Quattrocchio F, Wing JF, van der Woude K, Mol JNM, Koes R** (1998) Analysis of bHLH and MYB domain proteins: Species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. *Plant J* **13**: 475–488

- Ross J, Li Y, Lim E, Bowles DJ** (2001) Higher plant glycosyltransferases. *Genome Biol* **2**: 3004.3001–3004.3006
- Rumball W, Keogh RG, Miller JE, Claydon RB** (1997) ‘Grasslands ‘G27’’ red clover (*Trifolium pratense* L.) *NZ J Agricult Res* **40**: 369-372
- Sallaud C, El-Turk J, Bigarré L, Sevin H, Welle R, Esnault R** (1995) Nucleotide Sequences of Three Chalcone Reductase Genes from Alfalfa. *Plant Physiol* **108**: 869-870
- Sawada Y, Ayabe S** (2005) Multiple mutagenesis of P450 isoflavonoid synthase reveals a key active-site residue. *Biochem Biophys Res Commun* **330**: 907–913
- Sawada Y, Kinoshita K, Akashi T, Aoki T, Ayabe S** (2002) Key amino acid residues required for aryl migration catalysed by the cytochrome P450 2-hydroxyisoflavanone synthase. *Plant J* **31**: 555–564
- Schroder G, Wehinger E, Schroder J** (2002) Predicting the substrates of cloned plant O-methyltransferases. *Phytochem* **59**: 1–8
- Setchell KDR, Brown NM, Lydeking-Olsen E** (2002) The clinical importance of the metabolite equol—a clue to the effectiveness of soy and its isoflavones. *J Nutr* **132**: 3577–3584
- Shimada N, Aoki T, Sato S, Nakamura Y, Tabata S, Ayabe S** (2003) A Cluster of Genes Encodes the Two Types of Chalcone Isomerase Involved in the Biosynthesis of General Flavonoids and Legume-Specific 5-Deoxy(iso)flavonoids in *Lotus japonicus*. *Plant Physiol* **131**: 941–951
- Siminszky B, Corbin FT, Ward ER, Fleischmann TJ, Dewey RE** (1999) Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. *Proc Natl Acad Sci USA* **96**: 1750–1755
- Spaink HP** (2000) Root nodulation and infection factors produced by rhizobial bacteria. *Ann Rev Microbiol* **54**: 257–288

- Sreevidya VS, Rao CS, Sullia SB, Ladha JK, Reddy PM** (2006) Metabolic engineering of rice with soybean isoflavone synthase for promoting nodulation gene expression in rhizobia. *J Exp Bot* **57**: 1957–1969
- Stam M, Mol JN, Kooter JM** (1997) The Silence of Genes in Transgenic Plants. *Ann Bot* **79**: 3-12
- Steele CL, Gijzen M, Qutob D, Dixon RA** (1999) Molecular Characterization of the Enzyme Catalyzing the Aryl Migration Reaction of Isoflavonoid Biosynthesis in Soybean. *Arch Biochem Biophys* **367**: 146–150
- Subramanian S, Stacey G, Yu O** (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *Plant J* **48**: 261–273
- Subramanian S, Stacey G, Yu O** (2007) Distinct, crucial roles of flavonoids during legume nodulation. *Trends Plant Sci* **12**: 282-285
- Süß C, Hempel J, Zehner S, Krause A, Patschkowski T, Gottfert M** (2006) Identification of genistein-inducible and type III-secreted proteins of *Bradyrhizobium japonicum*. *J Biotechnol* **126**: 69-77
- Sutherland ORW, Russell GB, Biggs DR, Lane GA** (1980) Insect feeding deterrent activity of phytoalexin isoflavonoids. *Biochem Syst Ecol* **8**: 73-75
- Thomé OW** (1885) Flora von Deutschland, Österreich und der Schweiz - in Wort und Bild für Schule und Haus. Gera-Untermhaus.
- Thompson CJ, Movva NR, Tizard R, Crameri R, Davies JE, Lauwereys M, Botterman J** (1987) Characterization of the herbicide-resistance gene bar from *Streptomyces hygroscopicus*. *EMBO J* **6**: 2519–2523.
- Tian L, Dixon RA** (2006) Engineering isoflavone metabolism with an artificial bifunctional enzyme. *Planta* **224**: 496–507

Ulyatt MJ (1981) The feeding value of herbage: Can it be improved? *NZ Ag Sci* **15**: 200-205

van der Krol AR, Mur LA, Beld M, Mol JNM, Stuitje AR (1990) Flavonoid Genes in *Petunia*: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression. *The Plant Cell* **2**: 291-299

Van Etten HD, Matthews DE, Matthews PS (1989) Phytoalexin detoxification: Importance for pathogenicity and practical implications. *Ann Rev Phytopathol* **27**: 143-164

Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* **33**: 949-956

Voisey CR, White DWR, Dudas B, Ableby RD, Ealing PM, Scott AG (1994) *Agrobacterium*-mediated transformation of white clover using direct shoot organogenesis. *Plant Cell Rep* **13**: 309-314

Volpin H, Phillips DA, Okon Y, Kapulnik Y (1995) Suppression of an isoflavonoid phytoalexin defense response in mycorrhizal alfalfa roots. *Plant Physiol* **108**: 1449-1454

Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* **11**: 1337–1350

Wang LQ (2002) Mammalian phytoestrogens: Enterodiol and enterolactone. *J Chromatogr B* **777**: 289–309

Wang SF, Ridsdill-Smith TJ, Ghisalbert EL (1998) Role of isoflavonoids in resistance of subterranean clover trifoliate to the redlegged earth mite *Halotydeus destructor*. *J Chem Ecol* **24**: 2089-2100

Wang X (2010) Structure, function, and engineering of enzymes in isoflavonoid biosynthesis. *Funct Integr Genomics* DOI 10.1007/s10142-010-0197-9

- Warek U, Bedell JA, Budiman MA, Lakey N** (2005) GeneThresher® Methylation Filtering Technology – A Path to Rapid Gene and Marker Discovery in Plants, Czech J. Genet. Plant Breed **41**: 207. Proc. 5th International Triticeae Symposium, Prague, June 6–10, 2005
- Welle R, Grisebach H** (1988) Isolation of a novel NADPH-dependent reductase which coacts with chalcone synthase in the biosynthesis of 6'-deoxychalcone. FEBS Lett **236**: 221-225
- Welle R, Schroder G, Schiltz E, Grisebach H, Schroder J** (1991) Induced plant responses to pathogen attack. Analysis and heterologous expression of the key enzyme in the biosynthesis of phytoalexins in soybean (*Glycine max* L. Merr. cv. Harosoy 63). Eur J Biochem **196**: 423-430
- White DWR, Voisey C** (1994) Prolific Direct Plant-Regeneration From Cotyledons Of White Clover. Plant Cell Rep **13**: 303-308
- Winkel-Shirley B** (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiol **126**: 485–493
- Wojciechowski MF, Lavin M, Sanderson MJ** (2004) A phylogeny of legumes (Leguminosae) based on analysis of the plastid matk gene resolves many well-supported subclades within the family. Am J Bot **91**(11): 1846–1862
- Wojtaszek P, Stobiecki M, Gulewicz K** (1993) Role of Nitrogen and Plant Growth Regulators in the Exudation and Accumulation of Isoflavonoids by Roots of Intact White Lupin (*Lupinus albus* L.) Plants. J Plant Physiol **142**: 689-694
- Wroblewski T, Tomczak A, Micheltore R** (2005) Optimization of *Agrobacterium* mediated transient assays of gene expression in lettuce, tomato and *Arabidopsis*. Plant Biotechnol J **3**: 259–273
- Wu Q, Wang M, Simon JE** (2003) Determination of isoflavones in red clover and related species by high-performance liquid chromatography combined with ultraviolet and mass spectrometric detection. J Chromatogr A **1016**: 195-209

- Xie D-Y, Sharma S, Wright E, Wang Z-Y, Dixon RA** (2006) Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. *The Plant Journal* **45**: 895–907
- Yang H, Ahn JH, Ibrahim RK, Lee S, Lim Y** (2004) The three dimensional structure of *Arabidopsis thaliana* O-methyltransferase predicted by homology-based modeling. *J Mol Graph Model* **23**: 77–87
- Youn H, Lee S, Cho J, Oh H** (1991) *Insecticidal Isoflavone Glycoside from Maackia amurensis*. *Arch Pharm Res* **14**: 105-108
- Yu O, Jung W, Shi J, Croes RA, Fader GM, McGonigle B, Odell JT** (2000) Production of the Isoflavones Genistein and Daidzein in Non-Legume Dicot and Monocot Tissues. *Plant Phys* **124**: 781-793
- Yu O, McGonigle B** (2005) Metabolic engineering of isoflavone biosynthesis. In Sparks D, ed, *Advances in Agronomy Vol 86*, Elsevier, Newark, pp 147-190
- Yu O, Shi J, Hession AO, Maxwell CA, McGonigle B, Odell JT** (2003) Metabolic engineering to increase isoflavone biosynthesis in soybean seed. *Phytochem* **63**: 753–763
- Zubieta C, He X-Z, Dixon RA, Noel JP** (2001) Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nature Struct Biol* **8**: 271–279

Appendix 1

Plant tissue culture and bacterial culture media

Non selective tissue culture media contained no added antibiotic or herbicides. Selective tissue culture media contained 300mg L^{-1} TIM antibiotic, added when molten agar was below 60°C , and either 2.5mg L^{-1} AMM herbicide if the selection gene was BAR, or 200mg L^{-1} kanamycin (KAN) antibiotic if the selection gene was nptII.

Bacterial growth media contained $100\text{ }\mu\text{g ml}^{-1}$ ampicillin, $50\text{ }\mu\text{g ml}^{-1}$ kanamycin or $100\text{ }\mu\text{g ml}^{-1}$ spectinomycin, depending on which antibiotic resistance gene was present in the plasmid vector in the bacterium.

MS0 medium

1 x MS salts. 1 x MS vitamins, 30 g L^{-1} sucrose, 8.0 g L^{-1} agar. Solution was adjusted to pH 5.8 with potassium hydroxide and autoclaved.

MS101 medium

1 x MS salts. 1 x MS vitamins, 30 g L^{-1} sucrose, 8.0 g L^{-1} agar, 1.0 mg L^{-1} 6-benzylamino purine, 0.1 mg L^{-1} naphthalene acetic acid. Solution was adjusted to pH 5.8 with potassium hydroxide and autoclaved. All “MS” media and solutions are derived from Murashige and Skoog (1962)

CR7 medium

1 x MS salts, 1 x B5 vitamins, 30 g L^{-1} sucrose, 100 mg L^{-1} myo-inositol, 8 g L^{-1} Difco Agar, 1.0 mg L^{-1} 6-benzylamino purine, 0.05 mg L^{-1} 1-naphthaleneacetic acid. Solution was adjusted to pH 5.7 and autoclaved.

CR5 medium

1 x MS salts, 1 x B5 vitamins, 30 g L⁻¹ sucrose, 100 mg L⁻¹ myo-inositol, 8 g L⁻¹ Difco Agar, 6-benzylamino purine 0.1 mg L⁻¹, 1-naphthaleneacetic acid 0.05 mg L⁻¹. Solution was adjusted to pH 5.7 and autoclaved.

CR0 medium

1 x MS salts, 1 x B5 vitamins, 30 g L⁻¹ sucrose, 100 mg L⁻¹ myo-inositol, 8 g L⁻¹ Difco Agar. Solution was adjusted to pH 5.7 and autoclaved.

LB Medium

10 g L⁻¹ tryptone, 5 g L⁻¹ yeast extract, 10 g L⁻¹ NaCl. Adjust pH to 7.0. Sterilise by autoclaving. To make LB agar, add 15 g L⁻¹ agar. Sterilised by autoclaving.

SOC Medium

20 g L⁻¹ bacto tryptone, 5 g L⁻¹ bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose. Sterilised by autoclaving.

YM Medium

10 g L⁻¹ mannitol, 0.4 g L⁻¹ yeast extract, 0.5 g L⁻¹ K₂HPO₄·3H₂O, 0.1 g L⁻¹ NaCl, 0.2 g L⁻¹ MgSO₄·7H₂O, pH 7.0. For solid media add 15 g L⁻¹ agar. Sterilised by autoclaving.

MGL medium

5.0 g L⁻¹ mannitol, 1.0 g L⁻¹ L-glutamic acid, 250 mg L⁻¹ KH₂PO₄, 100 mg L⁻¹ MgSO₄, 100 mg L⁻¹ NaCl, 100 mg L⁻¹ biotin, 5.0 g L⁻¹ bacto tryptone, 2.5 g L⁻¹ yeast extract. Solution was adjusted to pH 7.0 with sodium hydroxide and autoclaved.

MS salts

1650 mg L⁻¹ NH₄NO₃, 440 mg L⁻¹ CaCl₂·H₂O, 370 mg L⁻¹ MgSO₄·7H₂O, 170 mg L⁻¹ KH₂PO₄, 1900 mg L⁻¹ KNO₃, 6.2 mg L⁻¹ H₃BO₃, 0.025 mg L⁻¹ CoCl₂·6H₂O, 0.025 mg L⁻¹ CuSO₄·5H₂O, 27.8 mg L⁻¹ FeSO₄·7H₂O, 22.3 mg L⁻¹ MnSO₄·4H₂O, 0.83 mg L⁻¹ KI, 0.25 mg L⁻¹ Na₂MoO₄·2H₂O, 8.6 mg L⁻¹ ZnSO₄·7H₂O, 37.3 mg L⁻¹ Na₂EDTA·2H₂O.

MS vitamins

Glycine 2 mg L⁻¹, *Myo*-Inositol 100 mg L⁻¹, Nicotinic Acid 0.5 mg L⁻¹, Pyridoxine-HCl 0.5 mg L⁻¹, Thiamine-HCl 0.1 mg L⁻¹.

B5 vitamins

Myo-Inositol 100 mg L⁻¹, Nicotinic Acid 1 mg L⁻¹, Pyridoxine-HCl 1 mg L⁻¹, Thiamine-HCl 10 mg L⁻¹.

1X TAE buffer

40 mM Tris-acetate and 2 mM Na₂EDTA, pH 8.5

10X loading buffer

50% glycerol + bromophenol blue dye.

1Kb+ ladder

10 µl 1Kb+ ladder (Invitrogen), 20 µl 10X loading buffer, 70 µl water

TE buffer

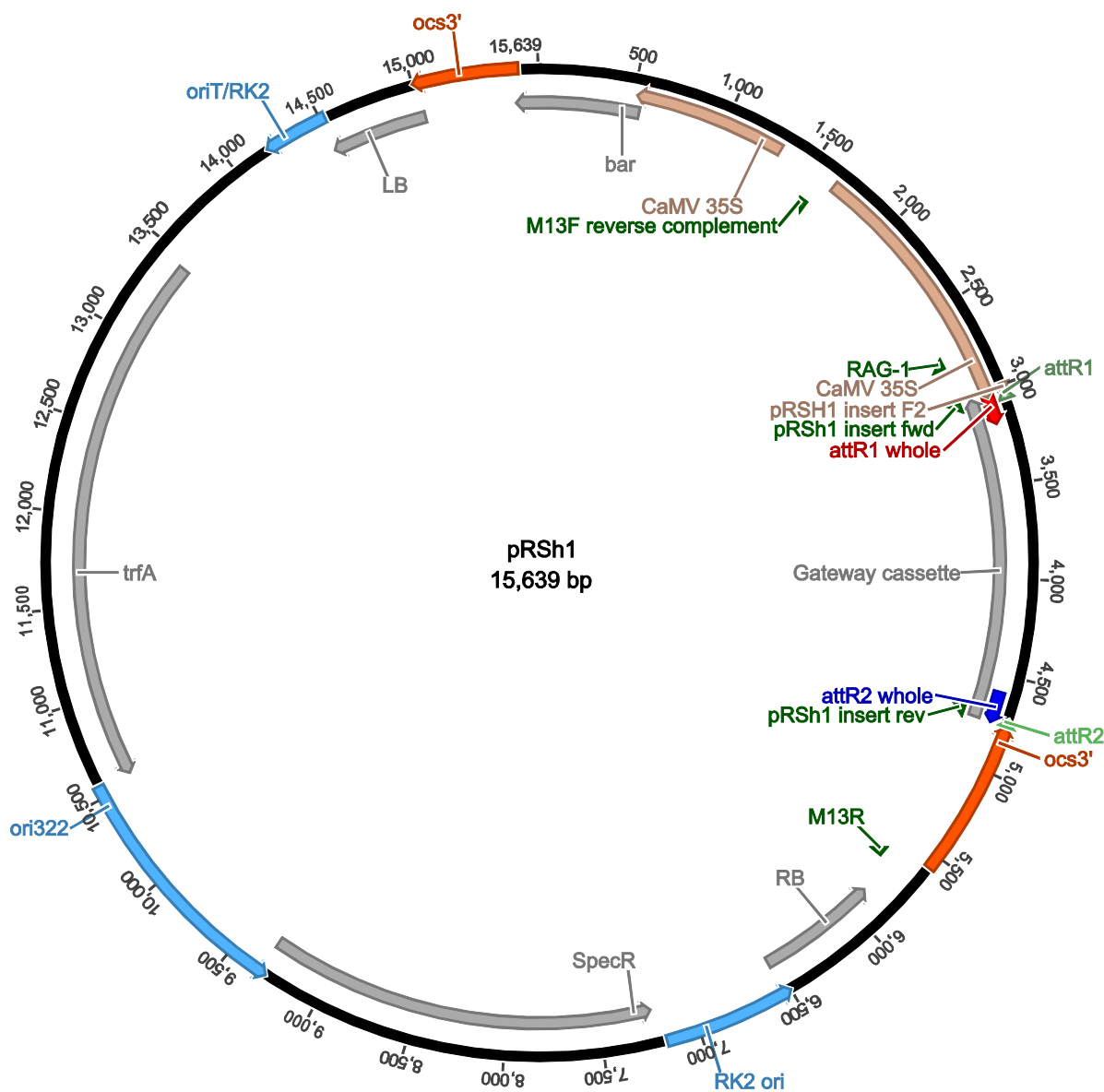
T₁₀E₁: 10 mM Tris-HCL, 1 mM Na₂EDTA.

T₁₀E_{0.1}: 10 mM Tris-HCL, 0.1mM Na₂EDTA

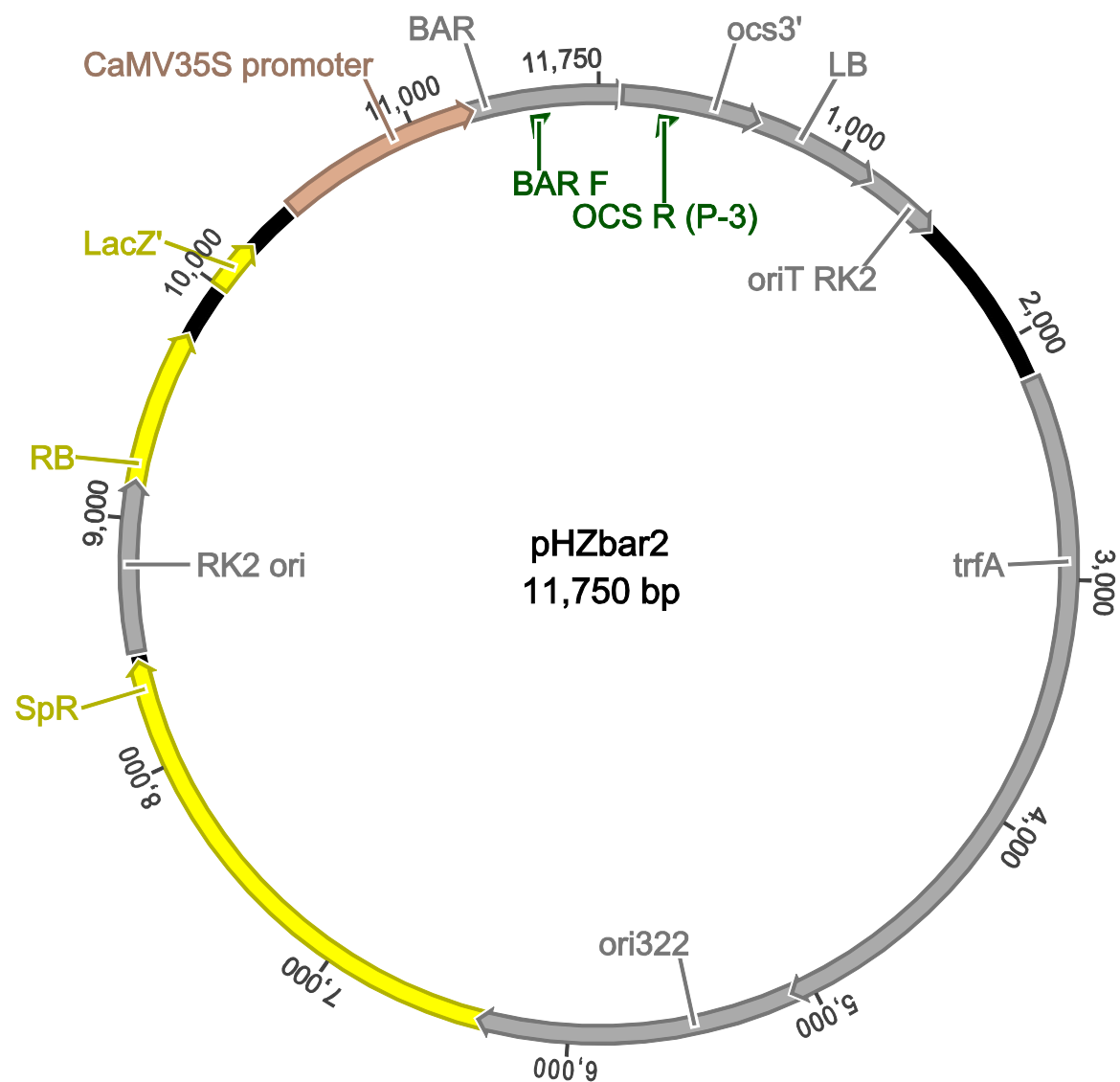
Appendix 2

Vector maps

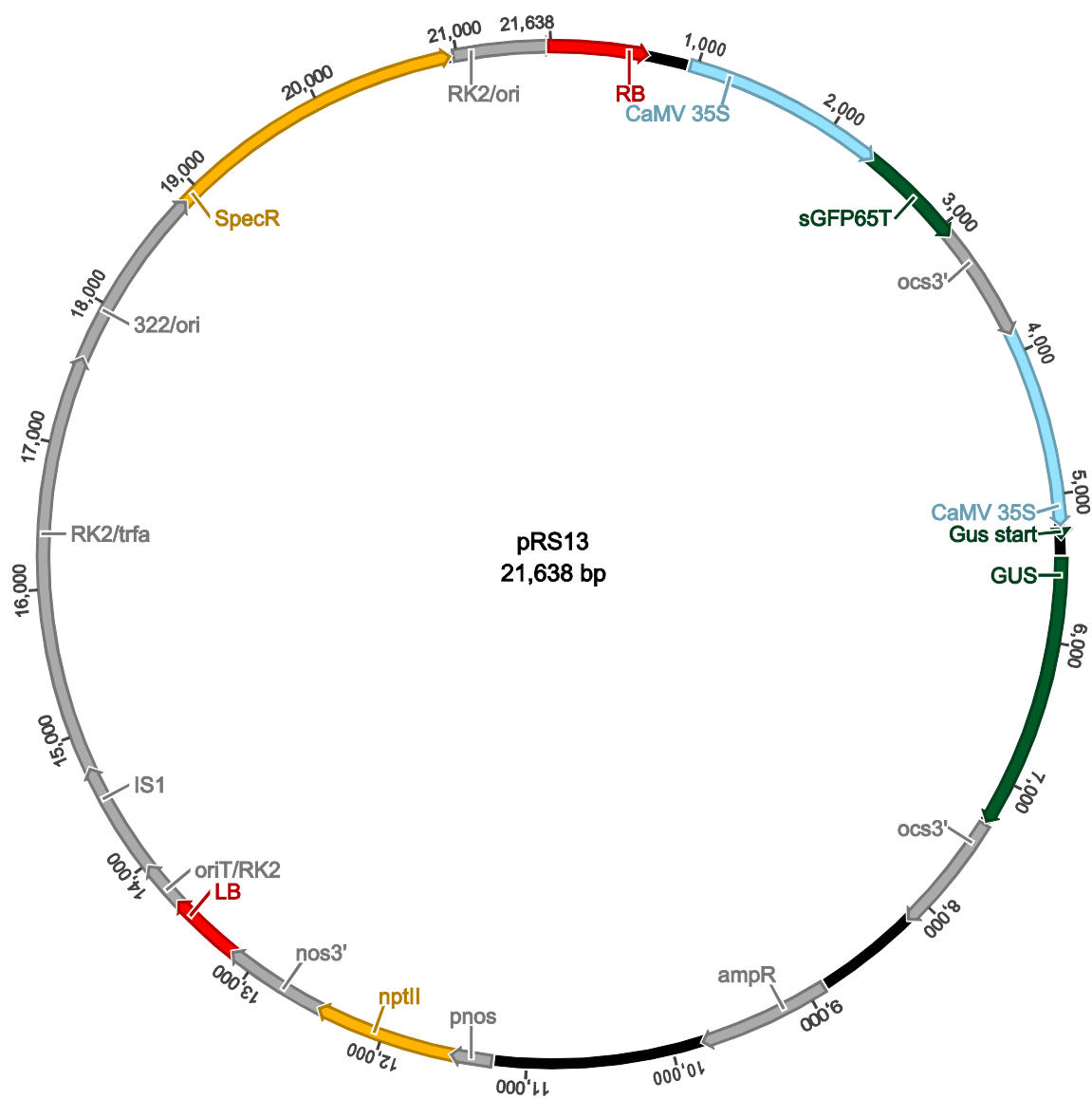
pRSh1



pHZbar



pRS13



Appendix 3

IFS2_12 sequence

The full length sequence of white clover IFS2_12 is listed below. The sequence of the single intron is underlined, the sequence before and after the intron are open reading frames one and two.

ATGTTGGTTGAACTTGCAGTTGCTTTACTAGTCATTGCTCTGTTTCATACATTTGCGTCCAACACCCACCGCTAA
ATCAAAAGCACTTCGCCATCTTCCAAATCCACCAAGTCCAAAACCTCGTCTTCATTTCATTGGTCATCTTCAT
CTTTTAGATAATCCTCTTCTTCACCACTCTTTAATCCGTTTAGGTGAACGTTATGGCCCTTTGTATTCTCTTTAC
TTTGGTTCCATGCCTACTGTTGTAGCTTCCACCCCTGATCTTTTCAAACCTCTTCTTCAAACCTCATGAAGCTAC
TTCTTTTAACACAAGGTTTCAAACCTCTGCTATTAGACGTTTAACTTATGATAACTCTGTTGCTATGGTTCCTT
TTGGACCTTACTGGAAGTTCATTAGAAAACCTATTATGAATGATCTTCTTAATGCTACCACTGTCAACAAGTT
GAGACCCTTGAGGAGCAAAGAAATCCATAAAGTTCTTAAGATTATGGCTAACAGTGCTGAAACTCAACAGCC
ACTTAATGTCAGTGAAGGAGCTTCTTAAATGGACAAATAGTACAATCTCAACCATGATGTTGGGTGAGGCTGA
AGAGGTTAGAGATATTGCTCGTGATGTTCTTAAGATCTTTGGGGAATATAGTATTATGGACTTCATTGGGCCT
TTGAAGATTTTAAAGAAGTTTGGGAACCTATGAGCAGAGGATTGATGCTATTTTAAACAAGTATGATCCTATCA
TTGAAAAGGTTATTAAGAAAAGACAAGAAATTTGTGAACAAAAGAAGAAGTGGAGAAATTCTAGAGGGTGA
GGAGGAGAGTGTAGTCTTCTTGATACTTTGCTTGAATATGCTCAGGATGAGACCATGGAGATCAAAATTAC
CAAGGAACAAATTAAGGGACTTGTGTGGTGAGTTCTTTTCTTCTAATAATTGCTTTGTTTATTGAATACAC
ATATTACATTTTAGGAAAAATCACTTCTACAAATCGTTCATTGTAAATTTAAATTTCTAACAACAACGTAAGT
GTTGTAAAGAAGTAGAAGTTAATAATTATCCGATTCTCAAAATTGAAATAAATTTTAAATTAATATTGTTTT
GATTACAATAACAAAGAATAAAGATATAACTCATATGGAACCAAAACAATATATCATTATTTTATTGTAAA
CACTGATTCACGATAAAGAACCTATAAAGGTGGTCCGCCTCCCCTATATTTTAGATAAAAATTGGCTCTCTGG
TTATATGCTAACTTAAATTTCTTTGTCTACAGGATTTCTTCTCTGCTGGGACAGACTCCACGGCTGTGGCAAC
AGAATGGACTTTGGCAGAGCTCATCAATAATCCGAGGGTGTGAAGAAAGCTCGGGAGGAGGTAGAATCTG
TTGTGGGAAAAGATAGACTTGTGACGAATCAGATGTTTCAAGATCTTCCCTACATTAGAGCCATGGTAAAGG
AGGCATTCCGCTTGACCCACCACTACCTGTGGTTAAGAGAAAATGTACCGAGGAGTGTGAGATCAACGGGT
ATGTGATTCCAGAGGGAGCATTGATACTTTTCAATGTTTGGCAAGTTGGAAGAGACCCAAAATATTGGGAGA
AGCCATTGGAATCCGTCCCGAGAGATTCTTAGAAAATCCTGGTGAAGGTGAAGCAGCTGCAGTTGATCTTA
GGGGTCAACATTTACACTTCTACCATTTGGGTCTGGAAGGAGAATGTGTCCTGGAGTTAATTTGGCTACAGC
AGGAATGGCCACACTACTTGCATCTATTATCCAGTGCTTTGATCTACAAGTGCCAGGTCCACATGGACAAAT
ATTGAAAGGTAGTGTGCTAAAGTTAGCATGGATGAGAGACCTGGGCTCACAGTTCCTAGAGCACATAATCT
TGTGTGTGTTCCACTTGCAAGAGCTGGCGTTGCATCTAAACTCCTTTCTTAA

Appendix 4

Primer sequences

name	sequence 5'-3'	Primer designed for/to:
ANT1_F+attB1	ggggacaagttgtacaaaaagcaggctatgaacagtacatctatg	AY348870 - tomato ANT1
ANT1_R+attB2	ggggaccactttgtacaagaaagctgggttaatcaagtagattcca	AY348870 - tomato ANT1
CHR_F_flanking	caccatcattcaacatagctcaaac	flanking primer
CHR_F_TrEST	tcaagtgaatgcctgtgg	internal primer
CHR_R_flanking	ttgcctaaggccaatagag	flanking primer
CHR_R_TrEST	attgagtcctggcttggtg	internal primer
CHR-CDS_GW_F	ggggacaagttgtacaaaaagcaggctaacaacatgggtagtggtgaaa	CHRg02 CDS
CHR-CDS_GW_R	ggggaccactttgtacaagaaagctgggttatattcatcatcccataggtca	CHRg02 CDS
IFS_F_flanking	caccggtgaaattgcagttgctttac	flanking primer
IFS_F_mRNAd	atgttgstkgaaamtgcattg	coding sequence primer
IFS_F_mRNAd+CACC	caccatgtgstkgaaamtgcattg	coding sequence primer
IFS_F_TrEST	atcaaaagcacttcgccatc	internal primer
IFS_R_flanking	tggcatcaacaacattcagc	flanking primer
IFS_R_mRNA1	ttaggacgaaatgagtttagcc	coding sequence primer
IFS_R_mRNA2	ttaagaaaggagtttagatgcaacg	coding sequence primer
IFS_R_TrEST	aattcttgcaagcggaaacac	internal primer
IFS_RT_F1	ctagagggtgaggaggagag	expressed sequence of IFS
IFS_RT_R1	gcagagaagaaatc[intron]cacaac	expressed sequence of IFS
IFS/CHI EcoR1 GW F	ggggacaagttgtacaaaaagcaggctgaattcatgttgctgaactgca	IFS/CHI fusion gene
IFS/CHI HindIII GW R	ggggaccactttgtacaagaaagctgggtaagcttcagttccaatcttgaa	IFS/CHI fusion gene
IOMT_F_flanking	caccaacataacaaaacacaaaccaac	flanking primer
IOMT_F_mRNAd	atgkcttcatcaatyaatggycg	coding sequence primer
IOMT_F_mRNAd+attB1	ggggacaagttgtacaaaaagcaggctatgkcttcatcaatyaatggycg	coding sequence primer
IOMT_F_mRNAd+CACC	caccatgkcttcatcaatyaatggycg	coding sequence primer
IOMT_F_TrEST	catcaatcaatggcgcacaa	internal primer
IOMT_R_flanking	aggatcaatggaaacatgacag	flanking primer
IOMT_R_mRNAd	ttawggatarakctcaataagagac	coding sequence primer
IOMT_R_mRNAd+attB2	ggggaccactttgtacaagaaagctgggttawggatarakctcaataagagac	coding sequence primer
IOMT_R_TrEST	ccttatgcaatccttatcaatcc	internal primer
pRSH1_F2	acgctcgaggaattcggtta	pRSH1 insert
IFS_5'_int_R	tgcttttgatttgagtg	IFS internal
P-3_OCS_rev	gagctacacatgctcagg	pRSh1
M13 F	caggaaacagctatgac	cloning vectors
M13 R	ctggccgctggtttac	cloning vectors